Single-cell atlas of pancreatic cancer reveals microenvironment ecosystems associated with prognosis and treatment responses

Constantinos Demetriou

University College Oxford

University of Oxford

Supervisor: Prof Shisong Jiang

Co-supervisor: Prof Eric O'Neill

A thesis presented for the degree of Doctor of Philosophy

Michaelmas Term 2025

Abstract

Due to its late-stage diagnosis and resistance to broad range of therapies, pancreatic ductal adenocarcinoma (PDAC) remains one of the most lethal solid malignancies. The tumour microenvironment (TME), comprising of diverse malignant, stromal, and immune cell states, exhibits significant heterogeneity which profoundly influences patient prognosis, metastatic potential, and treatment responses. Advances in single-cell RNA-sequencing (scRNA-seq) technologies have enabled studies to perform high-resolution characterization of the neoplastic and TME cellular landscape, overcoming previous limitations of bulk transcriptomic approaches. Yet, the limited number of donors, variable experimental techniques, and inconsistent cell type definitions prevent these studies from serving as a universal reference for this disease. Integrating multiple single-cell datasets to construct cellular atlases can address these limitations, enabling a holistic interrogation of the cellular landscape while capturing extensive inter-patient variability necessary for effective patient stratification.

Therefore, nine publicly available PDAC datasets were integrated to construct a comprehensive single-cell atlas. This single-cell atlas includes data from 126 individuals and a total of 565,584 cells, encompassing a range of disease states, and treatment conditions. We reveal a diverse repertoire of cell states, many of which are strongly prognostic with patient outcomes across independent cohorts, highlighting their potential as biomarkers for clinical prognosis. Importantly, we discover five robust cellular ecosystems (CEs) associated with distinct clinical outcomes, malignant cell states, and genomic alterations. Notably, one CE defined by immunosuppressive and proliferative cell states across multiple cell types significantly predicts response to immunotherapy yet exhibits marked resistant to chemotherapy across multiple cohorts. Additionally, the identified CEs show limited concordance with previously reported prognostic subtypes, indicating that our classification framework provides novel biological and clinical insights into the PDAC TME. Taken together, these CEs may facilitate effective patient stratification and may inform the development of novel therapeutic strategies.

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my supervisors, Prof Shisong Jiang, and Prof Eric O’Neill, for your immense support, encouragement, and continuous guidance throughout my DPhil.

To Prof Nicola Sibson, thank you for your invaluable advice and help during this journey. To Joanne Russell, thank you for your massive help and encouragement throughout my DPhil. I am also deeply grateful to the Medical Research Council for funding my DPhil.

I would also like to thank the members of the Jiang and O’Neill groups, past and present, for all the help and support I received during my time in the Department. I am grateful to Yuqian for teaching me various lab techniques and offering valuable guidance. To Mo and Ash, thank you for your encouragement, support, and for making the work environment so enjoyable.

To my family, thank you for all the love and support throughout the years. To Marinos and Demetris, thank you for the good times, the laughs and for always being there for me. And to Olga, this journey would not have been possible without your endless support, encouragement, and belief in me. For that, and so much more, thank you.

Publications

- Russell SN, **Demetriou C**, Valenzano G, Evans A, Go S, Stanly T, Hazini A, Willenbrock F, Gordon-Weeks AN, Mukherjee S, Tesson M, Morton JP, O'Neill E, Jones KI. Induction of macrophage efferocytosis in pancreatic cancer via PI3Kγ inhibition and radiotherapy promotes tumour control. Gut. 2025 Apr 7;74(5):825-839.
- Go S, **Demetriou C**, Valenzano G, Hughes S, Lanfredini S, Ferry H, Arbe-Barnes E, Sivakumar S, Bashford-Rogers R, Middleton MR, Mukherjee S, Morton J, Jones K, Neill EO. Tissue-resident natural killer cells support survival in pancreatic cancer through promotion of cDC1-CD8 T activity. Elife. 2024 Dec 10;13:RP92672.
- Miller, P., Akama-Garren, E.H., Owen, R.P. *et al.* p53 inhibitor iASPP is an unexpected suppressor of KRAS and inflammation-driven pancreatic cancer. *Cell Death Differ* 30, 1619–1635 (2023).

List of Contents

List of Figures and Tables

List of Abbreviations

5hmc	5-hydroxymethylcytosine
ADEX	Aberrantly differentiated endocrine exocrine
ADH-F	Adhesive
AE	Adaptive enriched
AMY2A	Amylase 2A
ApCAF	Antigen presenting CAF
APOE	Apolipoprotein E
ARI	Adjusted rand index
ASW	Average silhouette width
AUC	Area under the curve
BH	Benjamini-Hochberg
BICC1	Bicaudal C homolog 1
BRCA	Breast cancer gene
BRD4	Bromodomain-containing protein 4
BRPC	Borderline resectable pancreatic cancer
CA19-9	Carbohydrate 19-9
CAF	Cancer-associated fibroblast
CAR	Chimeric antigen receptor
CARD	Conditional autoregressive-based deconvolution
CCL	C-C motif chemokine ligand
cDC	Conventional dendritic cell
CDH1	Cadherin 1
CDKN2A	Cyclin dependent kinase inhibitor 2A
CE	Cellular ecosystem
CET	Cellular ecosystem type
CFTR	Cystic fibrosis transmembrane conductance regulator
CNA	Copy number alterations
CONKO-001	Charité Onkologie
cpm	Counts per million
CPTAC	Clinical Proteomic Tumour Analysis Consortium
CR	Complete response
csCAF	Complement-secreting CAF
CSF1	Colony stimulating factor 1
CTLA-4	Cytotoxic T-lymphocyte antigen 4
CXC	C-X-C motif chemokine ligand
DEG	Differentially expressed gene
DFS	Disease-free survival
dMMR	Mismatch repair deficiency
DNA	Deoxyribonucleic acid
E2	Oestradiol
ECM	Extra-cellular matrix
EGFR	Epidermal growth factor receptor
ESPAC-1	European Study Group for Pancreatic Cancer 1
FA	Fatty acid
FOLFIRINOX	Fluorouracil, Oxaliplatin, Irinotecan, Leucovorin
FXYD3	FXYD domain-containing ion transport regulator 3
GATA6	GATA binding protein 6

GemCap	Gemcitabine and capecitabine
GITSG	Gastrointestinal Tumour Study Group
GSEA	Gene set enrichment analysis
HRR	Homologous recombination repair
iCAF	Inflammatory CAF
ICB	Immune checkpoint blockade
ICGC	International Cancer Genome Consortium
IE/F	Immune enriched fibrotic
IL	Interleukin
IMM	Immunomodulatory
ITH	Intra tumour heterogeneity
KRT	Keratin
LAPC	Locally advanced pancreatic cancer
LCM	Laser capture microdissection
LGALS9	Galectin-9
LISI	Local Inverse Simpson’s Index
LPP	
L-R	Ligand-receptor
maxstat	Maximally selected rank statistics
MDSC	Myeloid-derived suppressor cell
ME	Myeloid enriched
MHC	Major histocompatibility complex
MP	Meta programme
MSI	Microsatellite instability
MSLN	Mesothelin
MT	Metastatic tissue
MUC1	Mucin 1
myCAF	Myofibroblastic CAF
MYO	Myofibroblastic
NAT	Normal pancreas or adjacent to tumour
Neo-CTx	Neoadjuvant chemotherapy
NET	Neutrophil extracellular traps
NFκB	Nuclear Factor κ B
NGS	Next generation sequencing
NRT	Neurotropic
OS	Overall survival
PAM	Partition around medoids
PanIN	Pancreatic intraepithelial neoplasia
PARP	Poly (ADP-Ribose) polymerase
PCA	Principal component analysis
PD	Progressive disease
PDAC	Pancreatic ductal adenocarcinoma
pDC	Plasmacytoid dendritic cell
PD-L1	Programmed death-ligand 1
PDO	Patient-derived organoid
PDX	Patient-derived xenografts
PFS	Progression-free survival
PICD	Phagocytosis-induced cell death
POLO	Pancreas Cancer Olaparib Ongoing
PR	Partial response
PREOPANC-1	Preoperative Radiochemotherapy for pancreatic cancer

PRODIGE-24	Partenariat de Recherche en Oncologie Digestive
PRSS	Protease serine
PT	Primary tumour
PVR	Poliovirus receptor
QC	Quality control
QMA	Quasi mesenchymal
RMSE	Root mean squared error
RNA	Ribonucleic acid
ROC	Receiver operating characteristic
S100A4	S100 calcium-binding protein A4
SBRT	Stereotactic body radiotherapy
scRNA-seq	Single cell RNA sequencing
SD	Stable disease
SLC4A4	Solute Carrier Family 4
SMAD4	Mothers against decapentaplegic homolog 4
snRNA-seq	Single nucleus RNA sequencing
ST	Spatial transcriptomics
stSME	Spatial morphological gene expression
TAM	Tumour-associated macrophage
TCGA	The Cancer Genome Atlas
TET2	Tet methylcytosine dioxygenase 2
TF	Transcription factor
TFF1	Trefoil factor 1
Tfh	T follicular helper
Th1	T helper 1
TIGIT	T cell immunoreceptor with Ig and ITIM domains
TME	Tumour microenvironment
TNFα	Tumour necrosis factor α
TP53	Tumour protein 53
TP63	Tumour protein 63
Treg	T regulatory cell
TREM2	Triggering receptor expressed on myeloid cells 2
trNK	Tissue-resident natural killer
TSG	Tumour suppressor gene
UMAP	Uniform manifold approximation and projection
WES	Whole exome sequencing
αSMA	Alpha smooth muscle actin

Chapter 1

Introduction

1.1 Challenges in the clinical management of PDAC

PDAC, which accounts for 90% of pancreatic cancers, is a highly lethal solid malignancy with rising cause of cancer-related deaths worldwide[1]. According to epidemiological studies, PDAC is currently the twelfth most common cancer, and the sixth leading cause of cancer-related deaths worldwide[2]. Additionally, due to its increasing annual incidence of 1%, PDAC is projected to become the second-leading cause of cancer-related mortality by 2030[3]. Growing prevalence of risk factors such as age, smoking, alcohol consumption, obesity, diabetes mellitus, and chronic pancreatitis, contribute to the increase in incidence of PDAC[4]. The poor prognosis of PDAC is largely attributed to the lack of effective screening methods and the late onset of symptoms. Specifically, only around 20% of patients present with resectable disease at the time of diagnosis, while most patients present locally advanced (30-35%) or metastatic (50-55%) disease[3]. Even after curative-intent surgery, the majority of patients develop disease recurrence with a five-year survival rate of 12-27%[5].

Over the past two decades, key clinical trials including the Charité Onkologie (CONKO-001) trial[6], and European Study Group for Pancreatic Cancer 4 (ESPAC-4)[7], have established the clinical benefit of adjuvant gemcitabine, as a monotherapy or in combination with capecitabine, over resection alone in patients with non-metastatic PDAC[8]. Moreover, the Partenariat de Recherche en Oncologie Digestive (PRODIGE-24) trial showed that FOLFIRINOX demonstrated improved overall survival (OS), and disease-free survival (DFS) rates compared to gemcitabine, with the expense of more toxic effects[9]. As a result, FOLFIRINOX is considered the standard adjuvant chemotherapy regimen for patients with improved performance status.

In recent years, neoadjuvant chemotherapy (neo-CTx) has gained focus due to its effectiveness in borderline resectable pancreatic cancer (BRPC), and locally advanced pancreatic cancer (LAPC)[10]. Benefits of neo-CTx include early eradication of micrometastases, downstaging, enhanced suitability of patients for surgery, and increased rates of margin negative (R0) resections[11] [12]. For example, the PREOPANC-1 trial for BRPC showed favourable outcomes to neoadjuvant gemcitabine-based chemoradiotherapy compared to surgery alone[13]. Moreover, patients with locally advanced or metastatic disease receive multiagent chemotherapy regimens, including FOLFIRINOX, GemCap, and irinotecan/fluorouracil[3]. Similar to the resectable PDAC, FOLFIRINOX has demonstrated improved outcomes compared to gemcitabine alone[14]. However, despite modest improvements with recent chemotherapeutic regimens, the overall prognosis of PDAC remains poor, with a five-year survival rate of 11%[15].

Nevertheless, due to the limited efficacy and considerable toxicity of systemic chemotherapy, targeted therapeutic approaches exploiting specific molecular vulnerabilities in PDAC have emerged as promising alternatives. For instance, PDAC patients harbouring germline or somatic *BRCA* mutations exhibit defective homologous recombination repair (HRR), requiring the Poly (ADP-Ribose) Polymerase (PARP) enzyme the repairs DNA single-strand breaks. Consequently, PARP inhibitors exploit synthetic lethality by selectively targeting HRR-deficient cancer cells while sparing normal cells with intact repair pathways. The Pancreas Cancer Olaparib Ongoing (POLO) study has shown that metastatic PDAC patients with germline BRCA mutations demonstrated improved progression-free survival (PFS) when treated with maintenance olaparib compared to placebo[16]. However, only a small fraction of PDAC patients is likely to benefit from this treatment, as germline *BRCA1/2* mutations occur in around 5 to 7% of cases[17]. Additionally, despite this selectivity, around 20% of BRCA-mutated patients display primary or acquired resistance due to the *BRCA* reversion mutations that restore HRR proficiency[18].

Immune checkpoint blockade (ICB) therapy has demonstrated remarkable responses in various other cancers; however, it has shown poor efficacy in unselected PDAC patients[19]. For instance, various phase II trials testing antibodies against cytotoxic T-lymphocyte antigen 4 (anti-CTLA-4) as a monotherapy or in combination with antibodies against programmed death-ligand 1 (anti-PDL-1) showed only modest survival benefits[20] [21]. Furthermore, combination of ICB with systemic chemotherapy, aimed at increasing neoantigen load and enhancing immune responses, has not resulted in improved responses[22]. Nevertheless, notable exception to ICB sensitivity is seen in patients harbouring microsatellite instability (MSI) or mismatch repair deficiency (dMMR). In the phase II KEYNOTE-158 trial, MSI-high (MSI-H) PDAC patients demonstrated durable responses to pembrolizumab monotherapy, albeit with lower efficacy compared to MSI-H colorectal cancer patients[23]. What is more, the extremely low prevalence of MSI-H PDAC patients, ranging from 0.8% to 2%, limits the broad accessibility of ICB in this disease[24].

Furthermore, despite the incredible success of chimeric antigen receptor (CAR) T cell therapy across cancer types[25], their efficacy in PDAC remains limited. For instance, several phase I trials evaluating mesothelin (MSLN)-directed CAR T cells have shown limited efficacy in PDAC[26] [27]. Similar to ICB, key challenges limiting CAR-T efficacy include extensive tumour heterogeneity, the dense fibrotic tumour microenvironment (TME) that restricts T cell infiltration, and immunesuppressive mechanisms that drive T cell exhaustion thereby impairing their function.

Despite advancements in chemotherapy, targeted therapies, and immunotherapy, treatment resistance remains a significant obstacle in PDAC, and survival outcomes remain dismal. A key contributor to poor treatment efficacy is the absence of robust predictive biomarkers that can guide patient stratification and improve therapeutic benefit. Thus, there is an urgent need to identify accurate molecular and TME biomarkers that can enable personalized treatment approaches, and improve efficacy while minimizing toxicity.

1.2 Tumour heterogeneity of PDAC

1.2.1 Genomic alterations

Advancements in next generation sequencing (NGS) have enabled a comprehensive characterization of the mutational landscape of PDAC, revealing a large number of mutations and somatic copy number alterations (CNAs)[28 29]. Activating mutations in Kirsten rat sarcoma (*KRAS)* are near ubiquitous (90% cases), along with deletions of key tumour suppressor genes (TSGs) Cyclin Dependent Kinase Inhibitor 2A (*CDKN2A),* Mothers against decapentaplegic homolog 4 (*SMAD4*)*,* and Tumour protein 53 (*TP53*) (50-75% cases)[29]. This is then followed by a long tail of less frequently mutated genes, involved in chromatin modification, DNA damage repair, and other mechanisms[29 30]. Furthermore, the process of malignant evolution has been extensively studied, and two models have been proposed. The stepwise progression model involves the gradual accumulation of the core driver mutations from pre-invasive pancreatic intraepithelial neoplasia (PanIN), towards infiltrating ductal carcinoma[31]. However, the more recently described punctuated model is based on the observation of simultaneous driver gene inactivations through chromosomal instability, leading to rapid tumour development and metastasis[32]. The latter model is more consistent with clinical findings of rapidly metastasizing tumours in certain patients (Fig. 1.1)[33].

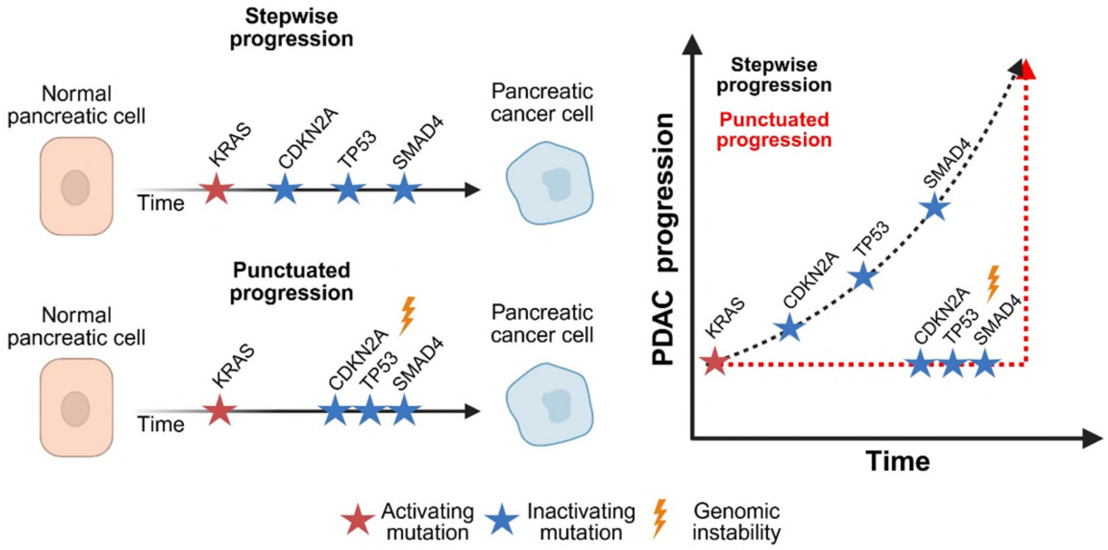


Figure 1.1 Stepwise and punctuated development models of PDAC. Proposed accumulation of somatic mutations in key driver genes (stepwise progression), while punctuated model involves accelerated progression during genomic instability of key driver genes. Created with Biorender.com

Despite their essential role for understanding PDAC carcinogenesis, these frequently mutated genes have not been able to stratify patients for therapeutic decisions[34]. For instance, while *KRAS* mutation status has been linked to patient prognosis, no clear associations with treatment response have been identified[35]. However, recent efforts have yielded promising small-molecule inhibitors targeting the most common $KRAS^{G12D}$ allele, including MRTX1133[36]. Yet, despite their strong preclinical activity, adaptive resistance limits the long-term clinical efficacy of these inhibitors, underscoring the need for novel combination strategies[37]. Conversely, the minority of PDAC tumours lacking *KRAS* mutations (<10%) represent a clinically distinct subgroup that are sensitive to targeted therapies, such as anti-epidermal growth factor receptor (anti-EGFR) nimotuzumab[38]. Collectively, the limited utility of genomic biomarkers in guiding treatment selection has shifted attention towards transcriptomic profiling.

1.2.2 PDAC transcriptomic subtypes

In recent years, large-scale gene expression studies of primary PDAC tumours have produced three clinically relevant classification systems published in rapid succession. These included the Collisson et al.[39] three-subtype classification (classical, quasi-mesenchymal (QMA), or exocrine-like), the Moffitt et al.[40] two- subtype classification (basal and classical), followed by the Bailey et al.[41] four-subtype classification (squamous, immunogenic, pancreatic progenitor or aberrantly differentiated endocrine exocrine (ADEX)). Based on their prognostic, and biological similarities, these classifications converged to two major subtypes, the basal-like (QMA, basal, squamous), and the classical/progenitor subtypes, with the former associated with poor differentiation, worse outcomes, and resistance to therapy[42]. Separate studies, including The Cancer Genome Atlas (TCGA), revealed that the immunogenic and ADEX subtypes from Bailey et al., and the exocrine-like subtype from Collisson et al., reflected contaminating gene expression of non-malignant cells, as they were mainly found in low-tumour cellularity samples[28].

Furthermore, Chan-Seng-Yue et al. performed *de novo* reclassification of PDAC subtypes in patients of primary and metastatic tumours, refining the two subtypes into five: “basal-like A”, “basal-like B”, “hybrid”, “classical A”, and “classical B” (Fig. 1.2)[42]. This analysis revealed that basal-like A were enriched in metastatic disease, and basal-like B tumours were more common in localized disease. Importantly, this study also identified previously understudied genomic alterations that could contribute to the divergence of the basal-like and classical phenotypes. Biallelic loss of *SMAD4* and GATA binding protein (*GATA6*) amplification were observed in classical phenotypes, while biallelic loss of *TP53* and/or *CDKN2A* along with *KRAS* mutant allele amplification were observed in basal-like phenotypes[42]. The increase in mutant *KRAS* gene dosage leading to more aggressive phenotype has been supported by previous mouse model and human studies[43] [35]. Thus, these findings posit that early acquisition of copy number alterations in driver gene mutations give rise to molecular subtypes, contributing to the substantial clinical heterogeneity of this aggressive disease.

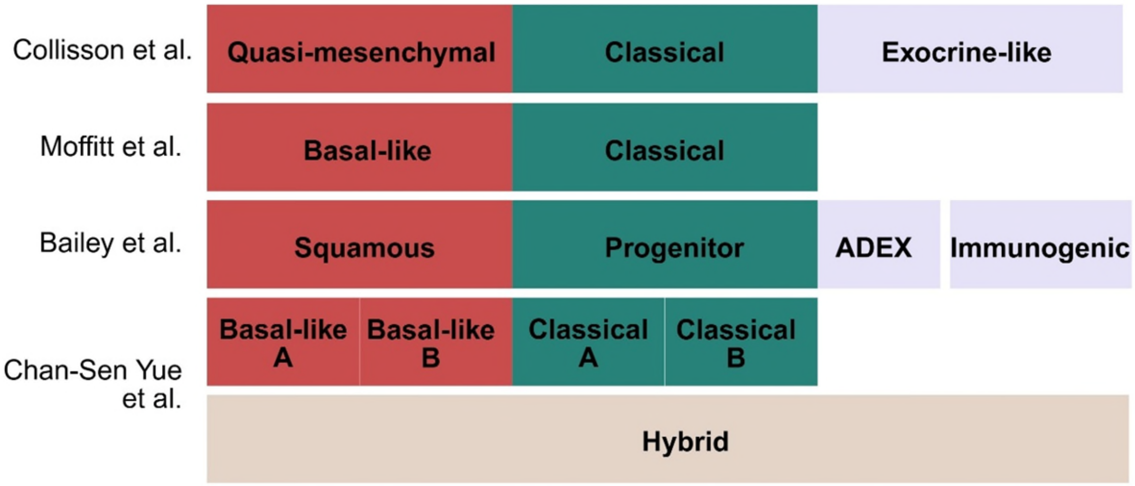


Figure 1.2 Previously published transcriptomic subtypes of PDAC. Each row indicates individual classification systems. Overlapped bars highlight the subtype similarities between classifications. The width of the bars does not indicate the frequency of subtypes. ADEX: Aberrantly differentiated endocrine exocrine. Created with Biorender.com

Beyond genomic alterations, other factors such as epigenetic transcriptional regulators, have been shown to shape PDAC subtype identity and display significant clinical implications. For instance, Eyres et al. revealed that aggressive basal-like PDAC subtypes arise from the loss of 5-hydroxymethylcytosine (5hmC), leading to epigenetic silencing of loci such as *GATA6*, which is critical for maintaining the differentiated classical phenotype[44]. Specifically, the loss of 5hmC occurs due to reduced expression of the 5-methylcytosine hydroxylase Tet methylcytosine dioxygenase 2 (TET2), which is itself driven by *SMAD4* loss. Importantly, pharmacologically stabilization of TET2 using metformin and vitamin C/ascorbic acid restored 5hmc and GATA6 levels, thereby re-establishing the classical phenotype, which is associated with improved survival. Similarly, another study using a PDX mouse model demostrated that pharmacological inhibition of the chromatin modifier Bromodomain-containing protein 4 (BRD4), which is involved in maintaining the aggressive basal-like subtype, promotes the classical phenotypic state[45]. These studies emphasize the potential of pharmacologically targeting subtype-specific chromatin modifiers and transcription factors (TFs) as a promising therapeutic strategy.

1.2.3 Subtype-specific biomarkers and treatment responses

The clinical utility of the PDAC subtypes for predicting treatment responses was first evaluated in the COMPASS (NCT02750657) trial[46]. This prospective study revealed that PDAC patients presenting the basal-like subtype, defined by low *in situ* ribonucleic acid (RNA) expression of *GATA6*, responded poorly to combination chemotherapy of GemCap with modified FOLFIRINOX (mFOLFIRINOX)[47]. Conversely, patients with high *GATA6* RNA expression showed significantly better responses, thus, proposing that GATA6 may act as a robust surrogate biomarker for subtype-based stratification and chemotherapy response. Likewise, basal-like specific markers have also been proposed, including Keratin 17 (KRT17)[48].

However, despite the promising potential of using surrogate markers for subtype stratification in clinical settings, advances in single-cell RNA-sequencing (scRNA-seq) have shown that these subtypes often co-exist within individual tumours[42]. This pronounced intra-tumour heterogeneity (ITH) poses a significant challenge, and is a key driver of therapy resistance, metastasis, and poor clinical outcomes. Additionally, recent studies revealed intermediate or “hybrid” transcriptional cell states expressing both GATA6 and KRT17[49], suggesting a continuum between classical and basal-like phenotypes, further complicating clinical decision making. Thus, understanding the intrinsic and extrinsic factors driving this subtype plasticity is key.

Furthermore, a recent study using pre-clinical models revealed that the local TME could modulate PDAC neoplastic subtypes[50]. Parallel heterogeneity across neoplastic, stromal, and immune cell states suggests that these compartments coordinate with each other during disease progression. However, the mechanisms underlying this cross-compartment heterogeneity, and its impact on prognosis and treatment response, remain poorly understood in PDAC, thus warranting further investigation.

1.3 Stromal and immune components of the PDAC TME

1.3.1 Cancer-associated fibroblast (CAF) heterogeneity

In recent years, scRNA-seq studies have expanded our understanding of the transcriptomic and phenotypic heterogeneity of CAFs in PDAC. CAFs have been shown to display diverse and context-dependent roles, exerting either tumour-restraining or tumour-promoting activities[51] [52] [53]. It is suggested that CAFs originate from tissue-resident stellate cells, which become activated during PDAC development[54].

Various studies have identified two broadly defined CAF subpopulations, referred to as myofibroblastic CAFs (myCAFs) and the inflammatory CAFs (iCAFs), that display distinct phenotypic, spatial, and functional activities[55] [56]. The myCAFs display upregulation of Transforming growth factor β (TGFβ) signalling, high expression of α-smooth muscle actin (αSMA) involved in enhanced tissue stiffness, and are localized adjacent to tumour regions. Targeted depletion of αSMA^+ CAFs in mouse models not only accelerated tumour growth but also enhanced the infiltration of immunosuppressive cells within the TME, suggesting a tumour-restraining role of myCAFs[52]. Conversely, iCAFs are defined by the expression of cytokines such as Interleukin-6 (IL-6), Leukaemia inhibitory factor (LIF), and C-X-C motif chemokine ligand 12

(CXCL12), and are found more distant to the neoplastic cells[51]. CAFs secreting CXCL12 and C-C motif chemokine ligand 2 (CCL2) have been shown to promote tumour progression and decreased overall survival, as these cytokines are involved with T cell exclusion, and recruitment of immunosuppressive myeloid cells, respectively[57] [58] [59]. Furthermore, Elyada et al. revealed a third CAF subpopulation termed antigen-presenting CAFs (apCAFs), defined by high expression of major histocompatibility complex II (MHCII) and CD74, however lack the expression of costimulatory molecules, indicating that apCAFs cannot function as expert antigen-presenting cells[56]. Instead, apCAFs have been shown to directly ligate and induce naïve $CD4^+$ T cells into regulatory T cells (Tregs), thus mediating immunosuppression[60].

Moreover, recent advancements of single-nucleus RNA-sequencing (snRNA-seq) enabled further expansion of CAF subsets, leading to four new classifications, including the adhesive (ADH-F), myofibroblastic (MYO), immunomodulatory (IMM), and neurotropic (NRT) programmes[61] . The MYO state resembled the myCAF phenotype, while the ADH-F, IMM, and NRT CAFs represented different iCAF subsets, however the apCAF was not identified in this study. Notably, this study revealed enrichment of the ADH-F and IMM CAF states following neo-CTx compared to untreated samples, indicating treatment-induced stromal cell remodelling[61]. Moreover, a separate scRNA-seq study identified a subset of CAFs expressing complement regulatory factors (*C3, C7,* and *CFD*), termed complement-secreting CAFs (csCAFs), which were shown to promote inflammation and immune activation[62]. This subset was enriched in stage I PDAC and decreased during tumour progression, suggesting a tumour-repressive role (Fig. 1.3)[63].

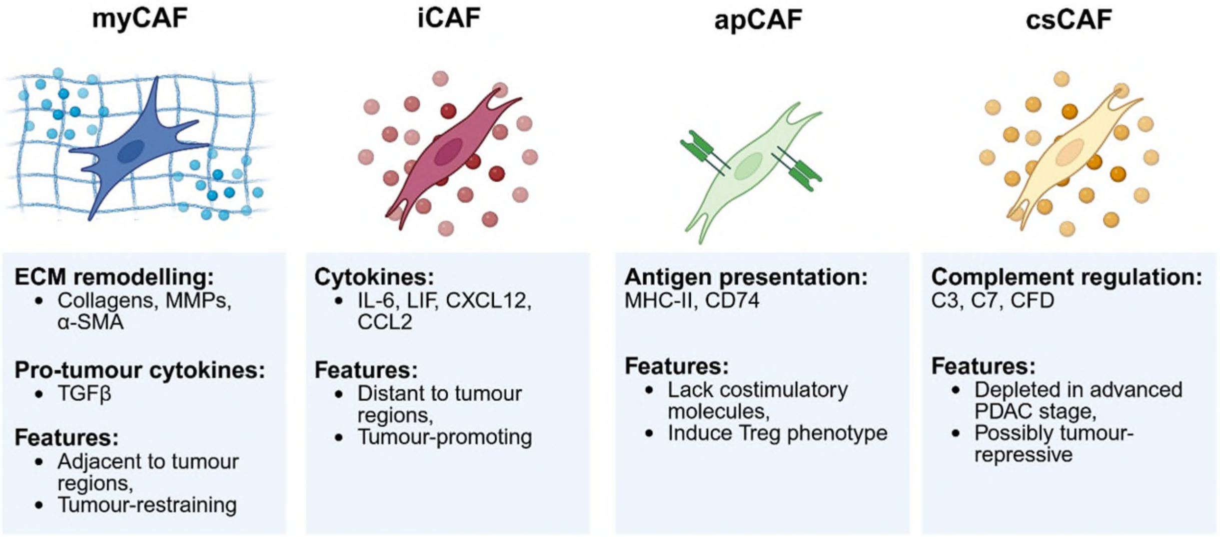


Figure 1.3 Key characteristics of previously identified CAF subtypes. Cytokines and other features are described from the corresponding publications. Created with Biorender.com

The identification of spatially distinct subTMEs revealed functional differences of CAFs that carry therapeutic and prognostic significance[64]. Specifically, CAFs originating from ‘reactive subTMEs’ were enriched for EMT, TGFβ signalling, as well as inflammatory signalling with immune hot features, whereas those derived from ‘deserted subTMEs’ displayed extra cellular matrix (ECM)-rich, lower CAF activation markers, and were enriched following chemotherapy. Reactive subTMEs were associated with the basal-like neoplastic phenotype, while deserted subTMEs were more associated with the classical phenotype, further supporting the notion that both the neoplastic and TME compartments coevolve during disease progression. However, conflicting evidence indicates that basal-like epithelium is associated with an ECM-rich stroma, while the immune-rich stroma tends to correlate with the classical epithelial phenotype[65]. Thus, no clear consensus between the CAF and PDAC neoplastic heterogeneity has been established, and given the emergence of newly defined CAF subsets, further investigation is warranted.

1.3.2 The immunosuppressive landscape of PDAC

Immunosuppression is a hallmark of PDAC, as its TME is predominantly enriched with immunosuppressive myeloid cells and characterized by extensive T cell dysfunction and exclusion mechanisms[66]. Additionally, PDAC is considered one of the most immunotherapy-resistant tumours, and even ICB response rates of MSI-H PDAC tumours are substantially worse compared to MSI-H tumours of other cancer types[67]. Consequently, a deeper understanding of the mechanisms underlying T cell immunosuppression, which limit the efficacy of immunotherapy, is urgently needed.

In recent years, scRNA-seq studies have revealed key communications that drive T cell immunosuppression in PDAC[68 69]. One key finding is the elevated expression of the immune checkpoint receptor T cell immunoreceptor with Ig and ITIM domains (*TIGIT*), which regulates NK and T cell-mediated recognition of cancer cells, and is strongly linked with T cell exhaustion[68]. Consistent with these observations, blockade of the CD155 (*PVR*)-TIGIT axis in combination with PD-1 co-inhibition and CD40 agonism, has been shown to reinvigorate tumour-reactive T cells in vivo[70]. Furthermore, an integrated analysis of the PDAC cellular communication networks identified Galectin-9 (encoded by *LGALS9*) as a primary immune checkpoint molecule derived from tumour-associated macrophages (TAMs)[71].

Furthermore, scRNA-seq studies have further enabled detailed characterization of the phenotypic and functional diversity of the PDAC immune landscape[68 69]. One study profiling the T cell landscape by scRNA-seq and T cell receptor-sequencing (TCR-seq) identified key CD8$^+$ T cell states, including the pre-dysfunctional (CD8-GZMK), exhausted (CD8T-CXCL13), tissue-resident memory (CD8T-ZNF683), and cytotoxic (CD8T-GZMB/PRF1) phenotypes[69]. Additionally, another study revealed activated Tregs expressing high levels of checkpoints *TIGIT, ICOS, CTLA4,* and *CD39*[72].

Moreover, multiple scRNA-seq studies have revealed extensive heterogeneity of the myeloid compartment in PDAC[71]. TAMs are characterized by high expression of complement factors, including *C1QA, C1QB,* and Triggering receptor expressed on myeloid cells 2 (*TREM2)*, which have been shown to drive T cell exhaustion[73]. Further studies revealed a subset TAMs expressing high apolipoprotein E (*APOE*), which has been reported to drive nuclear factor κ B (NFκB)-mediated immunosuppression by enhancing *CXCL1* and *CXCL5* expression in tumour cells, thereby recruiting immunosuppressive myeloid-derived suppressor cells (MDSCs) and suppressing CD8$^+$ T cell infiltration[74]. Additionally, a recent study identified IL-1β$^+$ TAMs associated with pathogenic inflammation and worse outcomes[75]. Taken together, these studies highlight the powerful utility of single-cell analyses in advancing our understanding of the immune cell landscape of PDAC, and elucidating key mechanisms of immunosuppression.

1.3.3 Cross-compartment communications of PDAC

Various studies have investigated the mechanisms by which TME extrinsic signals modulate PDAC subtype identity, and conversely, how distinct neoplastic cell states influence and potentially shape the surrounding TME. For example, infiltration of macrophages has been shown to maintain the basal-like phenotype in PDAC, and loss of TAMs through a colony stimulating factor 1 receptor (CSF1R) inhibitor caused a reduction of the squamous gene expression program and an increase of the ADEX and immunogenic phenotypes[76]. Moreover, Tu et al. revealed that TAM-derived TNF-a forces classical-like PDAC cells into the aggressive basal-like phenotype through a cJUN/AP1 lineage reprogramming[77]. In turn, basal-like cancer cells secrete CCL2, recruiting tumour necrosis factor α (TNF-α)-secreting macrophages, thus further maintaining basal-like subtype identity. Furthermore, experiments using orthotopic mouse models revealed that the basal-like/squamous tumour protein 63 (TP63) transcriptionally induces the expression of proinflammatory cytokines such as IL-1α and CXCL1 generating the iCAF phenotype, which in turn recruits immunosuppressive neutrophils[78]. Overall, these studies reveal clear bidirectional reprogramming between the neoplastic and immune and stromal cell compartments, which collectively shape the PDAC TME and impact prognosis and treatment responses.

However, the use of different approaches, such as bulk versus single-cell RNA-seq, and variations in cell type annotations across studies has led to diverse observations, and consequently, no consensus regarding the cross-compartment interactions has been reached. Furthermore, conflicting observations have been reported regarding tumour-TAM interactions. For instance, one study observed that classical PDAC tumours were enriched with angiogenic secreted phosphoprotein 1 (SPP1) $^{+}$ TAMs, whereas a recent scRNA-seq analysis associated this TAM population with basal-like tumours[79] [80]. Therefore, accurately characterizing the coordinated tumour-TME heterogeneity and the complex network of interactions is of considerable importance, as this may help elucidate mechanisms of treatment failure and guide effective patient stratification for prospective therapeutic strategies.

1.4 Single-cell atlas of PDAC

The application of scRNA-seq approaches has enabled detailed analyses of the tumour and TME cellular landscapes at high resolution. However, individual studies often exhibit specific biases due to differences in experimental protocols and sequencing technologies. Moreover, most datasets include a limited number of samples, lacking the scale and diversity required to fully capture the cellular and inter-patient heterogeneity of PDAC. To overcome these limitations, integration of individual studies to construct single-cell atlases has provided novel insights that were not accessible from individual datasets. For instance, reference atlases have enabled the discovery of unknown cell types[81] [82], the comparison of animal and human cell types[83], and patient stratification for disease subtypes[84]. Notably, a single-cell atlas of the human, which included 706 healthy samples across 35 tissues, revealed 12 cross-tissue coordinated multicellular ecosystems[85].

Therefore, to overcome the limitations of previous scRNA-seq studies of PDAC, we aim to construct a single-cell atlas to enable a holistic interrogation of the neoplastic and TME cell heterogeneity across a large cohort of patient samples. Our analysis will expand on previous efforts in several key aspects:

- Include normal pancreas or adjacent-to-tumour (NAT), primary PDAC tumour (PT), and metastatic tumour (MT) samples to investigate shifts in cell state abundance during disease progression.
- Include neo-CTx-treated samples to investigate chemotherapy-induced remodelling of cell state phenotypes and abundances.
- Utilize large scale single-cell atlas to identify robust PDAC cellular ecosystems (CEs) and investigate their associations with disease progression, patient prognosis, and treatment response.
- Perform *de novo* identification malignant cell states and investigate their relationships with the identified cellular ecosystems.
- Investigate genomic alterations to elucidate the molecular mechanisms underlying distinct cancer cell states, and cellular ecosystems.

Using these approaches, we constructed a large-scale single-cell atlas of PDAC comprising 565,584 cells from 126 individuals. Through this analysis, we revealed 53 TME and 6 malignant cell states, expanding on previous classifications and elucidating their associations with disease progression and treatment context. Furthermore, we identified five robust CEs, which were validated using spatial transcriptomics, bulk RNA-seq datasets, and cytokine profiling. Importantly, distinct cell states and CEs were significantly associated with patient outcomes, and exhibited differential responses to chemotherapy and ICB. Collectively, these findings indicate that the identified CEs may serve as promising biomarkers for patient stratification and novel treatment strategies for this highly aggressive disease.

Chapter 2

Methods

2.1 Curation of human PDAC scRNA-seq datasets

Publicly available scRNA-seq data of NAT, PT, and MT PDAC samples were collected from multiple sources. These include Steele et al.[68] (GSE155698), Zhang et al.[86] (GSE197177), Werba et al.[87] (GSE205013), Zhou et al.[88] (https://data.humantumoratlas.org), Lin et al.[89] (GSE154778), Moncada et al.[90] (GSE111672), Oh et al.[91] (GSE231535), Peng et al.[92] (GSA: CRA001160), and Xue et al.[93] (GSE202742).

2.2 Pre-processing and quality control of scRNA-seq datasets

The Seurat *(*v4.3.0) R package was used for main scRNA-seq processing and analysis steps. Datasets underwent pre-processing and quality control individually prior to integration. For each sample, low-quality cells with fewer than 200 genes and high proportion (>15%) of transcript counts derived from mitochondrial-encoded genes were removed. Removal of potential cell doublets was performed by excluding cells with extremely high read counts. Thresholds were determined in each study separately by visual inspection of the distributions, which are listed in Fig. 3.1b. Additionally, contaminating red blood cells with high proportion (>5%) of transcript counts derived from haemoglobin-encoded genes were removed. Available clinical information from each dataset, including tumour stage, gender, age, and tumour grade, were harmonized and included in Seurat meta-data. Processed datasets were then merged into a single Seurat file prior to integration.

2.3 Integration and cluster annotation of scRNA-seq datasets

The top 3000 highly variable protein coding genes were used for integration were selected using FindVariableFeatures, followed by SelectIntegrationFeatures. The merged dataset was converted to 'anndata' file format required for scVI integration using the scvi-tools python package[94]. An scVI model was created and trained using the recommended parameters[94]. Each patient sample was considered as an individual batch. The latent embeddings generated from scVI were projected to a two-dimensional space using the uniform manifold approximation and projection (UMAP) dimensionality reduction method[95]. Clusters were calculated using the Louvain algorithm at various resolutions (0.1-0.4) and cell types were annotated based on canonical cell marker expression derived from the literature. Next, cell type annotations were used from scVI integration to initialize the semi-supervised scANVI model with the recommended parameters from scvi-tools[96]. Similar to scVI, each patient sample was considered as an individual batch for scANVI integration. The latent embeddings generated from scANVI were projected to a two-dimensional space using the UMAP dimensionality reduction method. Clusters were calculated using the Louvain algorithm at various resolutions (0.1-0.4) and noisy/contaminating clusters were removed. Cell types were then re-annotated. The analysis workflow covering dataset quality control (QC), normalization and integration, is shown in Fig. 2.1.

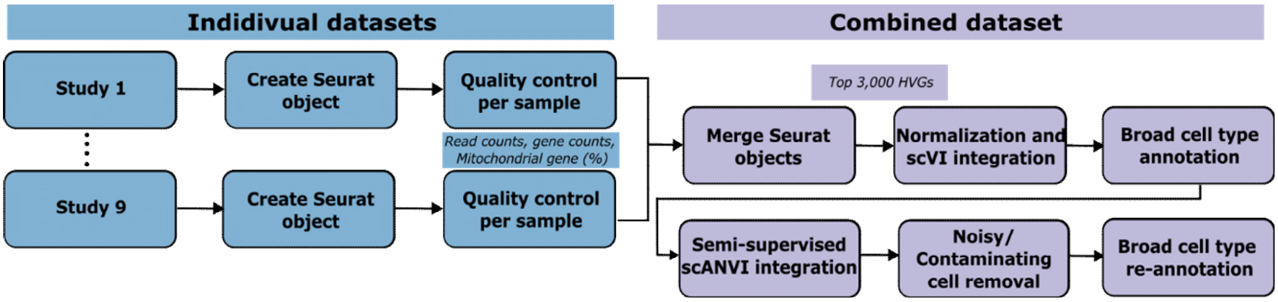


Figure 2.1 PDAC single cell atlas preparation workflow. Analysis workflow covering pre-processing, quality control, integration and cell type annotation of the PDAC single cell atlas. HVGs: highly-variable genes.

2.4 Benchmarking integration methods

To compare multiple integration methods, we down-sampled the PDAC single cell atlas and performed integration with FastMNN[97], Seurat's RPCA[98], Harmony[99], scVI[100], and scANVI[96] using default parameters. For both FastMNN, and Seurat's RPCA which require dataset splitting before integration, each study was used as an individual batch. Two metrics for both bio-conservation of cell type labels and batch correction of study labels were evaluated and are described below:

Bio-conservation metrics:

- Cell-type Local Inverse Simpson's Index (cLISI)[99]: Calculates the cell type mixing in local neighbourhoods from integrated graph embeddings. Because lower cLISI values indicated stronger separation of cell types, we inverted the metric (1 – mean cLISI) so that higher values corresponded to better cell type separation. The lisi (v1.0) R package was used for this calculation.

- Cell-type Adjusted Rand Index (ARI): Compares the overlap of two clustering labels[101]. Cell-type labels were compared with Louvain clustering computed on the integrated dataset using the same resolution (0.6) across integration methods. An ARI of 0 to 1 represents poor overlap and perfect match, respectively. The mclust (v6.1.1) R package was used to calculate ARI.

Batch correction metrics:

- Batch integration LISI (iLISI): Similar to cLISI but calculates batch mixing in local neighbourhoods from integrated graph embeddings. iLISI scores range from 1 to *N,* where *N* is the total number of batches (Study labels) in the dataset. Scores closer to *N* indicate better batch mixing.
- Batch integration ARI: Similar to cell-type ARI but compares the overlap of the Louvain clustering with the batch labels (Study), with values closer to 0 indicating better batch mixing. To ensure higher scores indicate better batch mixing, the scores were scaled by subtracting them from 1.

Next, average scores for bio-conservation and batch correction ranks were computed separately, and an overall integration score was computed by taking a weighted average of batch correction and bio-conservation scores, with a 40/60 weighting, respectively[102] [103].

2.5 Gene signature and pathway enrichment analysis

Curated gene signatures and their publication source are found in appendix (Supplementary Table 11). The Seurat's AddModuleScore function was used to calculate signature scores in individual cells. The scaled signatures scores were visualized in heatmap using the ComplexHeatmap (v2.24.1) R package. For pathway enrichment analysis, differentially expressed genes (DEGs) between groups were first calculated using the Seurat's FindAllMarkers function. Gene set enrichment analysis (GSEA) was then run using fgsea (v1.34.2) for MsigDB

Gene Ontology (GO): Biological Process (BP), and Hallmark gene signatures obtained from the msigdbr (v25.1.1) package[104]. All genes tested for differential expression was used for GSEA at 1000 permutations. The log fold change (logFC) was used as ranking metric.

2.6 Discovery of PDAC CEs

CEs were estimated using two complementary modules. Fist, cell state frequencies within their corresponding cell types were computed generating a normalized frequency matrix of 53 cell states (rows) x 129 patient samples (columns). Consensus clustering was then performed using ConsensusClusterPlus (v1.72.0) package to identify sample clusters with variable cell state frequencies. To ensure robustness, we ran 2 to 15 *k* clusters at 1000 permutations with the partitions around medoids (PAM) algorithm[105]. The optimal number of *k* clusters was selected based on the highest average silhouette width (ASW). Subsequently, a two-sided Wilcoxon test was performed to compare normalized cell state frequencies between consensus clusters. Cell states enriched within each consensus cluster were ranked based on a combined metric of logFC x -log10(p-value) values.

In the second approach, pairwise Pearson correlations were computed between all cell states using the normalized frequency matrix, generating a 53 x 53 cell state correlation matrix. Positively correlated cell state pairs ($R > 0.1$) were retained to construct a network graph using the igraph (v2.1.4) R package, with nodes representing cell states and edges weighted by correlation coefficients. The $R > 0.1$ threshold was applied as a lenient filter to remove spurious associations, while retaining weak to moderate cell state associations in sparse frequency data. Louvain clustering was applied on the network graph to detect communities of highly correlated cell states[106]. A resolution of 1 was performed.

Next, a Jaccard similarity matrix was generated to compare the consensus and Louvain clustering and overlapping cell states between both clustering approaches were considered as core CEs. The convergence between independent modules mitigates the limitations of using a lenient correlation threshold. Additional cell states that uniquely correlated with CE abundance were incorporated to the core CEs, generating the extended CEs. The igraph package was then used to visualize CE networks, with nodes coloured by cell type and the edges coloured by correlation coefficients. Analysis workflow is shown in Fig. 2.2.

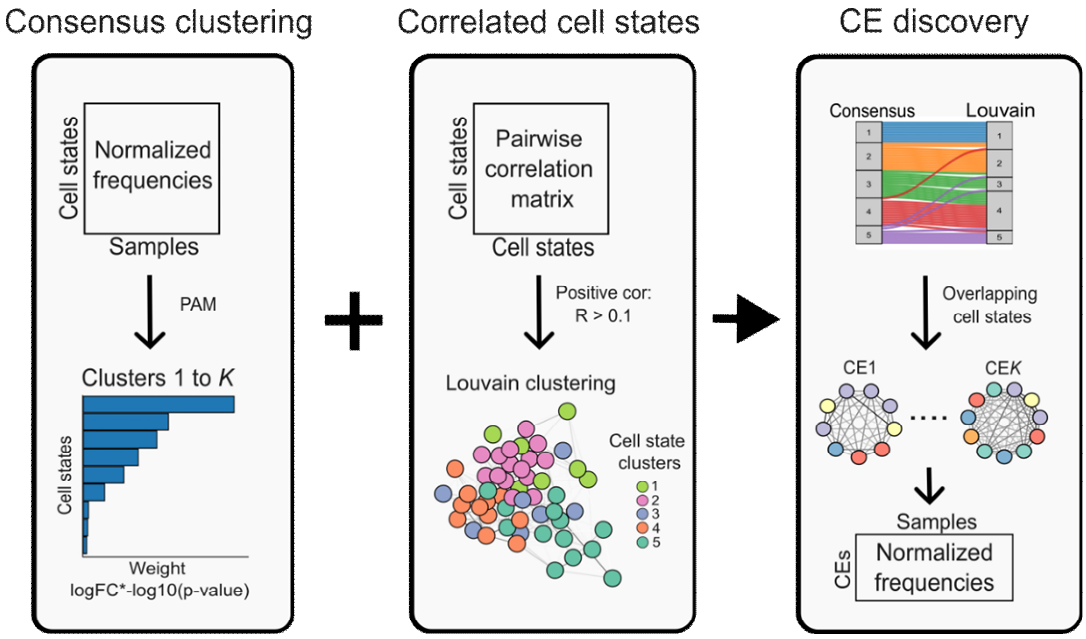


Figure 2.2 CE discovery workflow. First analysis involves consensus clustering of the cell state x sample normalized frequency matrix using PAM clustering. Cell states of clusters 1 to *K* are prioritized based on ranking metric (logFC * -log10(p-value)). Next analysis involves constructing a cell state x cell state pairwise correlation matrix and subsequent Louvain community clustering. The overlapping cell states between consensus and Louvain correlation clusters are considered as CEs. PAM: Partition around medoids; CE: Cellular ecosystem.

2.7 CE type (CET) classification of patient samples

The CE abundance was measured by taking the average of cell state frequencies within each CE, and a min-max normalization was applied so that CE abundances ranged from 0 to 1 in each sample. Next, each sample was assigned a CET label based on its most abundant CE. For example, if a sample exhibited the highest CE1 abundance across the five CEs, it was labelled as CET1.

2.8 Spatial transcriptomics (ST) analysis

2.8.1 Curation of ST data

Spatially resolved transcriptomics data were collected from Zhou et al.[107]. Data included 15 slide sections, with certain sections originating from the same patient at different regions.

2.8.2 Identification of cell states and CEs

For deconvolution of spatial transcriptomic data, we utilized conditional autoregressive-based deconvolution (CARD)[108], a non-negative matrix factorization model that uses cell type-specific gene expression information from scRNA-seq data as well as spatial correlation for accurate and robust deconvolution. We used the scRNA-seq integrated atlas as a reference to obtain cell state signatures. The scRNA-seq data was first subsampled to 100 cells per cell state, for all 53 TME cell states. By following the recommended parameters of CARD, we set $N = 5$ as the minimum number of counts for each gene, and $N = 5$ as the minimum number of counts for each spatial location. This resulted in the expected cell state abundance in each spot. Similar to patient samples (Methods 2.7), the CE abundance in each spot was computed by taking the average of cell state frequencies within each CE, and a min-max normalization was applied so that CE abundances ranged from 0 to 1 in each spot.

A unique feature of CARD is its ability to construct a refined tissue map with higher resolution than that measured in the original study[108]. We set $N = 10{,}000$ as the number of grids and used this feature for downstream analysis.

2.8.3 Colocalization analysis

For each CE, we calculated Pearson correlation coefficients between cell state frequencies using the refined frequency matrix. We compared correlations for pairs in which both cell states belonged to the same CE (“same CE: yes”) with those where only one cell state was part of the CE (“same CE: no”). For each pair type, the median correlation coefficient was computed across all spatial sections. Colocalization specificity between pair types was then assessed using a paired Wilcoxon signed-rank test. Global colocalization scores were computed as described previously[85]. Specifically, we defined a score for each CE as the proportion of correlations among “same CE: no” CE pairs that were less than or equal to the median correlation of “same CE: yes” CE pairs. This provided a normalized measure of colocalization relative to global correlations.

2.8.4 Spatial aggregation analysis

To evaluate the spatial aggregation of cell state components within CEs, we applied the local Moran’s Index (I)[109], using the spdep R package (v1.3-13). Spot coordinates of the refined frequency matrix were extracted using the CARD package and used to define spatial neighbourhoods via k-nearest neighbours ($k = 6$), which were then used to construct a spatial weights matrix. For each spot, local Moran’s I was calculated for each cell state as a measure of spatial autocorrelation and Pearson correlation coefficients between cell state local Moran’s I were computed. Next, similar to the colocalization analysis, correlations were compared between “same CE: yes” and “same CE: no” pair types, and an aggregation score was derived for each CE, using the local Moran’s I instead of estimated abundance.

2.8.5 Comparison of CEs with morphological-based clusters

To independently validate CEs, we applied the Spatial Morphological gene Expression (stSME) clustering method from the stLearn python package[110], which integrates gene expression, tissue morphology, and spatial proximity to adjust gene expression and improve cell type classification. The stSME algorithm was run with default parameters to generate an adjusted gene expression matrix for each spatial section. Louvain clustering implemented in the Seurat package was then performed on the stSME-normalized expression matrix to identify morphology-informed clusters. Validation was assessed by measuring the overlap between annotated CEs and these clusters, as well as by comparing pathway activity profiles.

2.9 Cell-cell communication analysis

2.9.1 Cell type communication analysis

Comparison of cell type communications across tissue types and treatments was performed using the multinichenetr R package (v2.1.0)[111], using the study labels as individual samples.

2.9.2 CE communication analysis

Given the large number of cells, we subsampled the single cell atlas to balance the contribution of each cell state. Specifically, we subsampled to 1,000 cells per cell state. In cases where cell states did not exceed 1,000 cells, all cells were included. This approach prevented bias towards cell states with larger cell numbers. Cell-cell interactions within CEs were then inferred using the CellChat (v2.2.0) package[112], by averaging results across the corresponding cell states. To compare pathway activities between CEs, we computed the average communication probability for each pathway, generating a matrix with communication pathways as rows and CEs as columns. For each CE, we retained the top 15 pathways from the ‘Secreted Signaling’ and ‘Cell-Cell Contact’ categories. Pathway communication probabilities were min-max normalized within each CE to range from 0 to 1.

2.9.3 CE-specific ligand-receptor pair analysis

Using the down-sampled atlas, we performed differential expression analysis for each cell type across CEs. To evaluate communication pathways, we utilized the CellChat database of ligand-receptor (L-R) pairs. A pathway was considered enriched in each CE if both the ligand and its corresponding receptor were significantly upregulated (LogFC > 0.1 and padj < 0.05) compared with the same cell type in other CEs. For example, if CE1 myeloid cells upregulated *CD86* while CE1 lymphoid cells upregulated *CTLA4* relative to their counterparts in other CEs, the *CD86-CTLA4* signalling pathway was considered specifically enriched in CE1.

2.10 Bulk RNA-seq data analysis

2.10.1 Curation of human bulk RNA-seq datasets

The PDAC datasets collected include TCGA pancreatic adenocarcinoma[113] (TCGAbiolinks[114] R package, v2.26.0), Clinical Proteomic Tumour Analysis Consortium (CPTAC)[115], (https://portal.gdc.cancer.gov/analysis_page?app=Projects), International Cancer Genome Consortium (ICGC) (https://dcc.icgc.org/), and the Puleo et al.[116] cohort (https://www.ebi.ac.uk/biostudies/arrayexpress/studies/E-MTAB-6134). The Puleo et al. cohort was split into two separate cohorts based on their sequencing location (France (Puleo FRA) and Belgium (Puleo BEL)). Available clinical data were collected from the original publications and harmonized (Appendix Table 12). Only patient samples with PDAC histology were included for downstream analyses. The liver-lung metastatic PDAC dataset was collected form Link et al[117] (GSE281129).

The neo-CTx chemotherapy dataset was obtained from Zhou et al[49]. The adjuvant CTx (adj-CTx) Oxford-Kidani microarray cohort was obtained in-house. Scoring of pre-processed IHC stained images from the in-house cohort are described in Diana et al.[118]. Combined classifications derived from the binary (low/high) scoring of a-SMA and CD68 were used for the analysis, generating four categories (a-SMA^{low}CD68^{low}, a-SMA^{high}CD68^{low}, a-SMA^{low}CD68^{high}, and a-SMA^{high}CD68^{high}).

Gene expression and clinical information from the immunotherapy cohorts were obtained from the Tumour Immunotherapy Gene Expression Resource (http://tiger.canceromics.org/) database. Gene expression and clinical information from the PDAC immunotherapy cohort were obtained from Parikh et al.[119]

To ensure consistency in our analyses, all datasets that were obtained as raw read counts were normalized using counts-per-million (cpm) with the addition of a pseudocount of 1. The neo-CTx (Zhou et al.), and in-house (Oxford-Kidani) cohorts were obtained in pre-normalized form.

2.10.2 Development of CTcaller and CScaller deconvolution methods

Differential expression analysis between cell types was performed in the down-sampled atlas and the top 100 genes (logFC > 0.5, padj < 0.05) were retained. For each cell type, the top n candidate genes whose expression positively correlated with its ground truth proportions (Spearman's correlation > 0.3, p < 0.05) were selected. A range of n values were tested (8-50), and the value that maximized the median Spearman's correlation across pooled cell types was chosen as optimal. This analysis was rerun with the optimal *n* to generate refined signature gene sets. These gene signatures were subsequently applied to bulk data by calculating the average score for each cell type in each sample. The average scores were then transformed into probability distributions using the softmax function[120], yielding normalized cell type abundances between 0 and 1 for each bulk sample.

2.10.3 Benchmarking of cellular deconvolution methods

The down-sampled atlas was used to construct a scRNA-seq reference matrix for deconvolution. The deconvolution tools used include CIBERSORTx[121], InstaPrism[122] (v0.1.6), MuSiC[123] (v1.0.0), and BisqueRNA[124] (reference-based method, v1.0.5) with their default parameters. The pseudo-bulk data (genes x sample matrix) was used as input to estimate cell type abundances across deconvolution methods. The estimated abundances were then compared to the ground-truth abundances. Pearson's and Spearman's correlation were computed using the cor() function from stats R package (v4.5.0) and the root mean squared error (RMSE) was calculated using the Metrics R package (v0.1.4). Benchmarking of deconvolution tools was repeated using a publicly available snRNA-seq dataset from Hwang et al.[61], using their pre-annotated labels as input.

2.10.4 Estimation of CE abundances in bulk data

To minimize skewness in cell type abundances caused by dominant cell type in certain samples, CE estimations were performed in two steps. First, cell type abundances in each sample were estimated using the CTcaller. Second, for each cell type, cell state proportions were calculated using the CScaller and softmax normalized to sum to 1. These proportions were then multiplied by the corresponding cell type abundance, so that the combined frequencies of all cell states across all cell types summed to 1. This approach produced a normalized frequency matrix that reduced skewing by dominant cell types while emphasizing within-cell-type heterogeneity. Next, CE abundance was measured by taking the average of cell state frequencies within each CE, and a softmax normalization was applied so that CE abundances ranged from 0 to 1 in each sample. Finally, each sample was assigned a CET label based on its most abundant CE, similar to the single-cell atlas (Methods 2.7).

2.10.5 Survival analysis of cell state and CE abundances

Multivariate Cox proportional hazards regression of the cell states and CE abundances was performed using the coxph function of the survival R package (v.3.8-3). Patient age, TNM stage, and gender were used as control variables. For TCGA, radiation therapy was also included as a control variable. Univariate Kaplan-Meier survival curves were plotted using the ggsurvplot function in the survminer R package (v0.4.9). Continuous variables, such as cell state and CE abundances were assigned to ‘low’ and ‘high’ groups based on the maximally selected rank statistics (maxstat) method calculated using the maxstat R package (v0.7-26).

2.10.6 CE-based cytokine profiles

The pre-normalized cytokine profiles of the five patient samples were obtained from the O’Neill group. The cytokine platform used was the Human Cytokine Array 6 from RayBio® (https://www.raybiotech.com/human-cytokine-array-g6-aah-cyt-g6). Cytokine protein symbols were converted to gene symbols. Next, differential expression analysis of the down-sampled single-cell atlas was performed using the Seurat FindAllMarkers function. Genes present in the cytokine profile array were used to filter the DEGs (logFC > 0.1 and padj < 0.05) across CEs.

2.11 Tumour cell analysis of the single-cell atlas

2.11.1 Identification of tumour cells

Among the epithelial cell clusters, malignant cells were distinguished from non-malignant epithelial cells by inferring CNAs for each cell using the copykat R package (v1.1.0) as described[125]. For each patient sample, chromosomal arms were defined as amplified or deleted if the scaled CNA score was above 0.1, and below -0.1, respectively. To account for tumour cells with low CNA burden, we calculated tumour and non-malignant epithelial gene signatures for each cell[107] (Supplementary Table 12). Epithelial cells with a malignant score above 0.1 and a normal-epithelial score below 0.1 were defined as tumour cells (low CNA).

2.11.2 Discovery of tumour cell states

Cancer cell states of malignant cells were identified as described previously[126]. For each individual sample with more than 100 tumour cells, clusters were calculated using the Seurat R package at four resolutions (0.4, 0.6, 0.8, 1). For each cluster, the top 50 DEGs were identified and only clusters with more than 10 cells and more than 20 DEGs were retained. These DEGs were defined as individual gene signatures. Within each sample, pairwise comparison of gene signatures was performed to remove redundant gene signatures. For each pair with Jaccard similarity > 0.75, the gene signature with fewer genes was removed. Consensus clustering of the Jaccard similarity between gene signatures across all tumours was performed. The ConsensusClusterPlus R package was run from 2 to 15 *k* clusters at 1000 permutations with the PAM algorithm. The optimal number of *k* clusters was selected based on the ASW. Next, the top 30 over-represented genes for each consensus cluster were obtained based on the logFC metric, and were considered as the cancer cell state gene signatures.

Each cell was assigned to a cell state with the highest normalized enrichment score if the second highest was no more than 95% of the highest score, otherwise the cell was considered 'unresolved'. DEGs were then computed between the annotated malignant cell states and underwent GSEA analysis using fgsea package for MsigDB Hallmark, and curated PDAC gene signatures. The logFC was used as ranking metric.

2.12 Genomic alteration analysis

Non-synonymous somatic mutation annotations (single nucleotide variants and small insertions and deletions) and discretized CNA (deep deletions, shallow loss, gains, and amplifications) values were obtained from both TCGA and CPTAC cohorts. Only the TCGA cohort included chromosomal arm discretized CNA values. Non-synonymous mutation differences between CETs were evaluated using the maftools R package (v2.22.20). Germline mutations, including *BRCA1/*2 mutations were not included in the analysis. Chromosomal arm and gene level CNA differences were compared between CETs using the Fisher's exact test from the stats R package (v.4.5.0). *P* values were adjusted using the Benjamini-Hochberg correction. Curated oncogenes and TSGs were obtained from the OncoKB database[127]. Distance-based phylogenetic analysis of the broad CNAs was performed using the ape R package (v5.8-1).

2.13 Statistics and reproducibility

Statistical analyses were performed in R (v.4.5.0). Comparisons of categorical groups were conducted using χ^2 tests or Fisher's exact test. Wilcoxon rank-sum tests were used for continuous variables and paired Wilcoxon signed-rank tests were used for paired comparisons. Log-rank tests were performed for survival analyses. Pearson and Spearman correlation coefficient was used when comparing two continuous variables. Receiver operating characteristic (ROC) curves for prediction of immunotherapy response were run using the pROC R package (v1.19.0.1). *P* values were corrected for multiple-hypothesis testing using the Benjamini-Hochberg (BH) procedure. Statistical significance was assumed at $P < 0.05$.

2.14 Code availability

The code used to generate the thesis results is available from the author upon reasonable request.

Chapter 3

Constructing the single cell atlas of PDAC

3.1 Introduction and Aims

Advancements in scRNA-seq technologies have allowed researchers to decode the heterogeneity of PDAC TME at a cellular level[128]. However, these studies are limited to small number of patient samples and their variable experimental protocols, sequencing technologies, and downstream analyses leading to study-specific biases and inconsistent cell type classifications.

To address this, integrated single-cell atlases capture extensive diversity across studies and serve as universal references for multiple cancer types[129 130]. Although Chijimatsu and others have established a PDAC reference atlas, these efforts were based predominantly on resected primary tumours[131]. However, integration of recently published PDAC datasets that include metastatic or neoadjuvant-treated samples remain limited, leaving important clinical contexts under-represented[132 133 107].

Moreover, the choice of integration method is paramount for correctly removing technical effects whilst preserving biologically relevant information. Common tools like Harmony[99] and Seurat's RPCA[98], have been shown to underperform in atlas-level datasets[102 134], and have failed to distinguish rare cell types[134].

Therefore, in this chapter we aim to construct a single cell atlas of pancreatic cancer with newly published datasets that include NAT, PT, and MT samples as well as patients treated with neo-

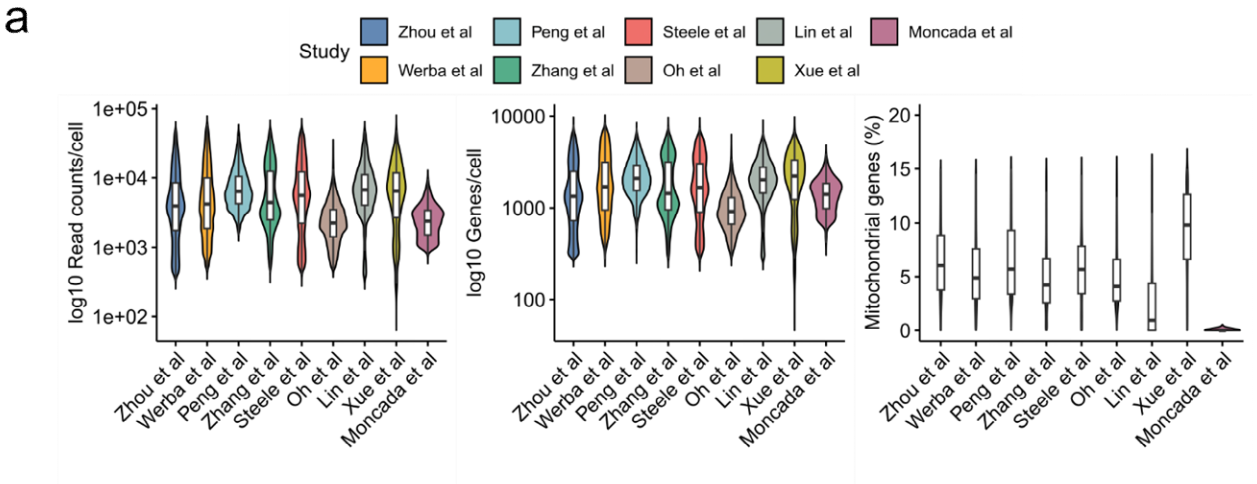
CTx. This atlas will cover extensive patient variability not captured in individual studies, and provide new insights into the PDAC TME. The specific aims of this chapter include:

- Perform consistent preprocessing and quality control of individual datasets to ensure compatibility prior to integration.
- Benchmark multiple integration methods using established metrics for bio-conservation and batch correction and use the top-performing method.
- Identify and annotate major cell types following integration, and investigate their associations with clinical features.
- Perform sub-clustering of major cell types to identify and characterise TME cell states
- Investigate changes in abundance and communication networks across disease progression and treatment conditions.

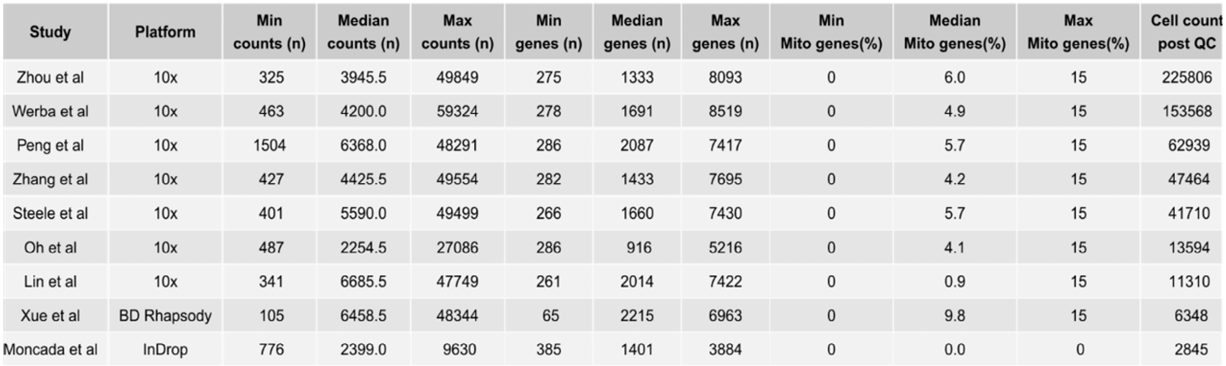
3.2 Quality control of scRNA-seq datasets

First for each sample, genes expressed in minimum of three cells were filtered out as this filtering threshold is commonly used in the single-cell community[135]. Cell QC was then performed based on three variables: (a) the number of reads per cell (read depth), (b) the number of genes detected per cell, and (c) and the fraction of counts detected from mitochondrial genes[135]. Cells with low count depth and low number of genes were excluded as they represent dying cells, and cells whose membrane is damaged leading to cytoplasmic mRNA leakage[135]. Cells with high mitochondrial counts were also excluded, as this indicates cell membrane damage[135].

Moreover, cells with contaminating red bloods cells, with high proportion of transcript counts derived from haemoglobin-encoded genes were excluded. Lastly, cells with high counts and number of genes were also excluded as they represent potential cell doublets[135]. The distributions of read counts, number of genes, and mitochondrial gene fractions following QC were comparable across studies (Fig. 3.1a). Samples from the majority of studies were sequenced using the 10x genomics platform, whereas samples from Xue et al. and Moncada et al., were sequenced using BD Rhapsody and InDrop systems, respectively (Supplementary Table 1). Summary statistics of these metrics, and the number of cells post QC for each study are shown in Fig. 3.1b.



b



Study	Platform	Min counts (n)	Median counts (n)	Max counts (n)	Min genes (n)	Median genes (n)	Max genes (n)	Min Mito genes(%)	Median Mito genes(%)	Max Mito genes(%)	Cell count post QC
Zhou et al	10x	325	3945.5	49849	275	1333	8093	0	6.0	15	225806
Werba et al	10x	463	4200.0	59324	278	1691	8519	0	4.9	15	153568
Peng et al	10x	1504	6368.0	48291	286	2087	7417	0	5.7	15	62939
Zhang et al	10x	427	4425.5	49554	282	1433	7695	0	4.2	15	47464
Steele et al	10x	401	5590.0	49499	266	1660	7430	0	5.7	15	41710
Oh et al	10x	487	2254.5	27086	286	916	5216	0	4.1	15	13594
Lin et al	10x	341	6685.5	47749	261	2014	7422	0	0.9	15	11310
Xue et al	BD Rhapsody	105	6458.5	48344	65	2215	6963	0	9.8	15	6348
Moncada et al	InDrop	776	2399.0	9630	385	1401	3884	0	0.0	0	2845

Figure 3.1 Quality control of PDAC scRNA-seq datasets. (a) Violin plots showing the distribution of read counts (top), number of genes (middle), and mitochondrial genes (%) (bottom) across studies. **(b)** Table displaying the summary statistics of read counts, gene counts, and mitochondrial gene fractions for each study.

3.3 Benchmarking integration tool performance

To optimally remove dataset-specific batch effects, we compared five integration methods using two metrics for both bio-conservation and batch correction: graph-based LISI which measures the label (cell type or batch) mixing in local neighbourhoods from integrated graph embeddings, and ARI which compares the overlap of Louvain-based clustering and labels (cell type or batch)[102] (Methods 2.4, Fig. 3.2a). To account for memory, we down-sampled the dataset to around 90,000 cells (16% of entire dataset) for benchmarking of integration methods. Harmony excelled in batch correction ranking first for graph integration iLISI and cluster/study ARI but underperfomed in both bio-conservation metrics, consistent with previous reports[136] (Fig. 3.2b). Moreover, scANVI was the top-performing tool for bio-conservation ranking first for graph cell-type cLISI and cluster/cell-type ARI and moderately performing for batch correction.

Next, an overall integration score was computed by taking a weighted average of batch correction and bio-conservation, with a 40/60 weighting, respectively[102]. scANVI ranked first, followed by scVI, in line with previous atlas-level studies[129] [136] (Fig. 3.2b). Therefore, scANVI was used to create an integrated embedding of 565,584 high-quality cells for all downstream analyses.

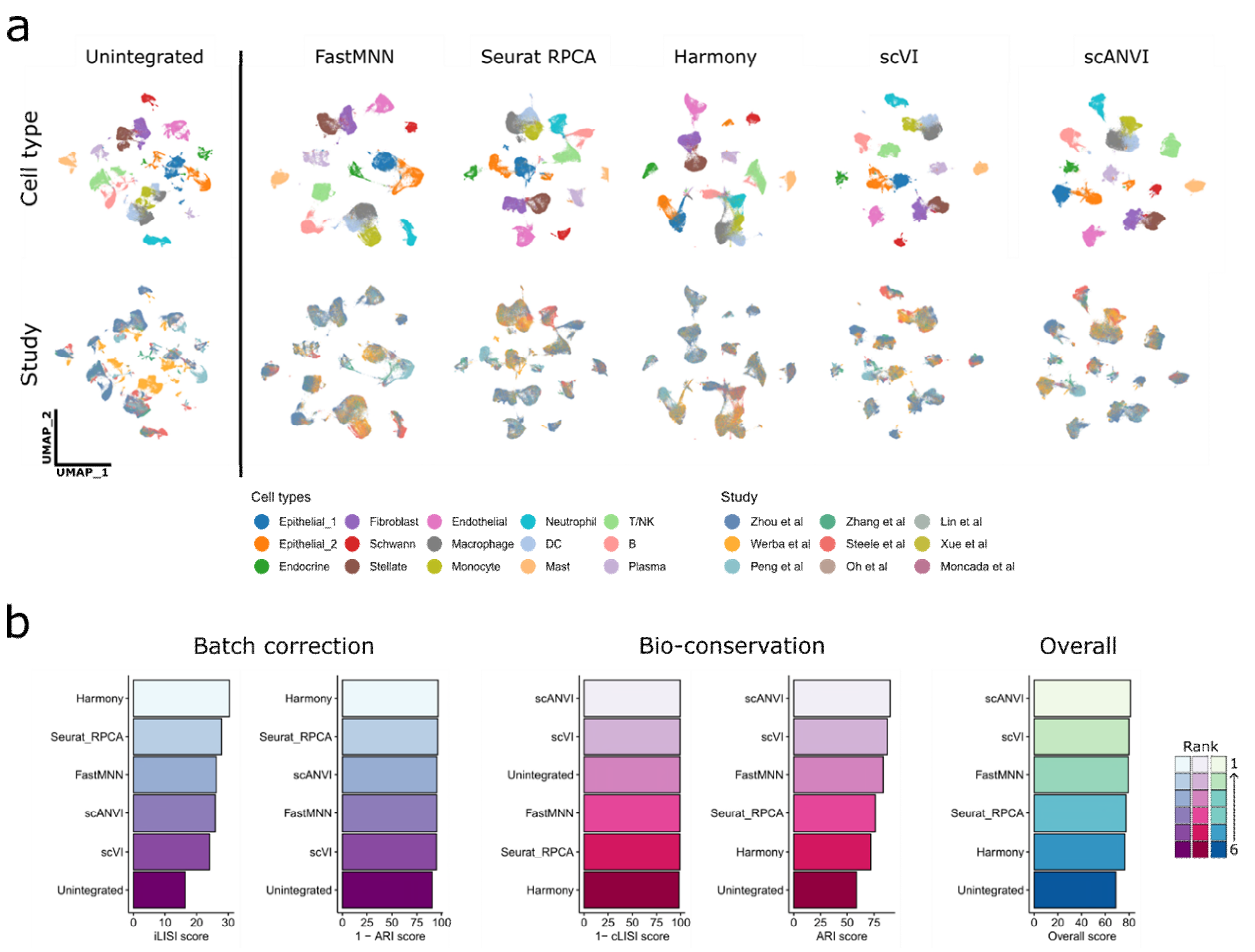


Figure 3.2 Comparison of integration tools in down-sampled dataset. (a) UMAP visualizations of PDAC scRNA-seq data following integration by different integration tools. Cells are coloured by cell type to show impact of integration on bio-conservation (top row), while cells are coloured by scRNA-seq study to show impact of integration on batch correction (bottom row). **(b)** Bar plots showing the performance metrics for batch correction (left) and bio-conservation (middle). Bars are coloured based on performance ranking. Overall score of the weighted average batch correction and bio-conservation scores (right). LISI: Local Inverse Sampson's Index; ARI: Adjusted Rand Index.

3.4 Broad cellular landscape of PDAC atlas

Following quality control and integration, a high quality dataset was constructed consisting of a 126 individuals from NAT, PT, MT samples, and samples treated with neo-CTx (n = 18) (Fig. 3.3a-b). Additionally, these datasets included patients with a diversity in TNM stage, sex, histological grade, and age, thus providing comprehensive patient variability. After cluster annotation, 15 major cell types were identified based on canonical cell marker expression (Fig. 3.3c-d). Broad cell type composition and clinical characteristics of samples is shown in Fig. 3.3e.

There were stark differences in cell type compositions across studies such as high T/NK abundance in Steele et al, Werba et al, and Zhou et al (Fig. 3.3e). Notably, Steele et al. showed the highest abundance of neutrophils. These study-specific differences underscore the importance of dataset integration to account for variations in sample preparation and processing.

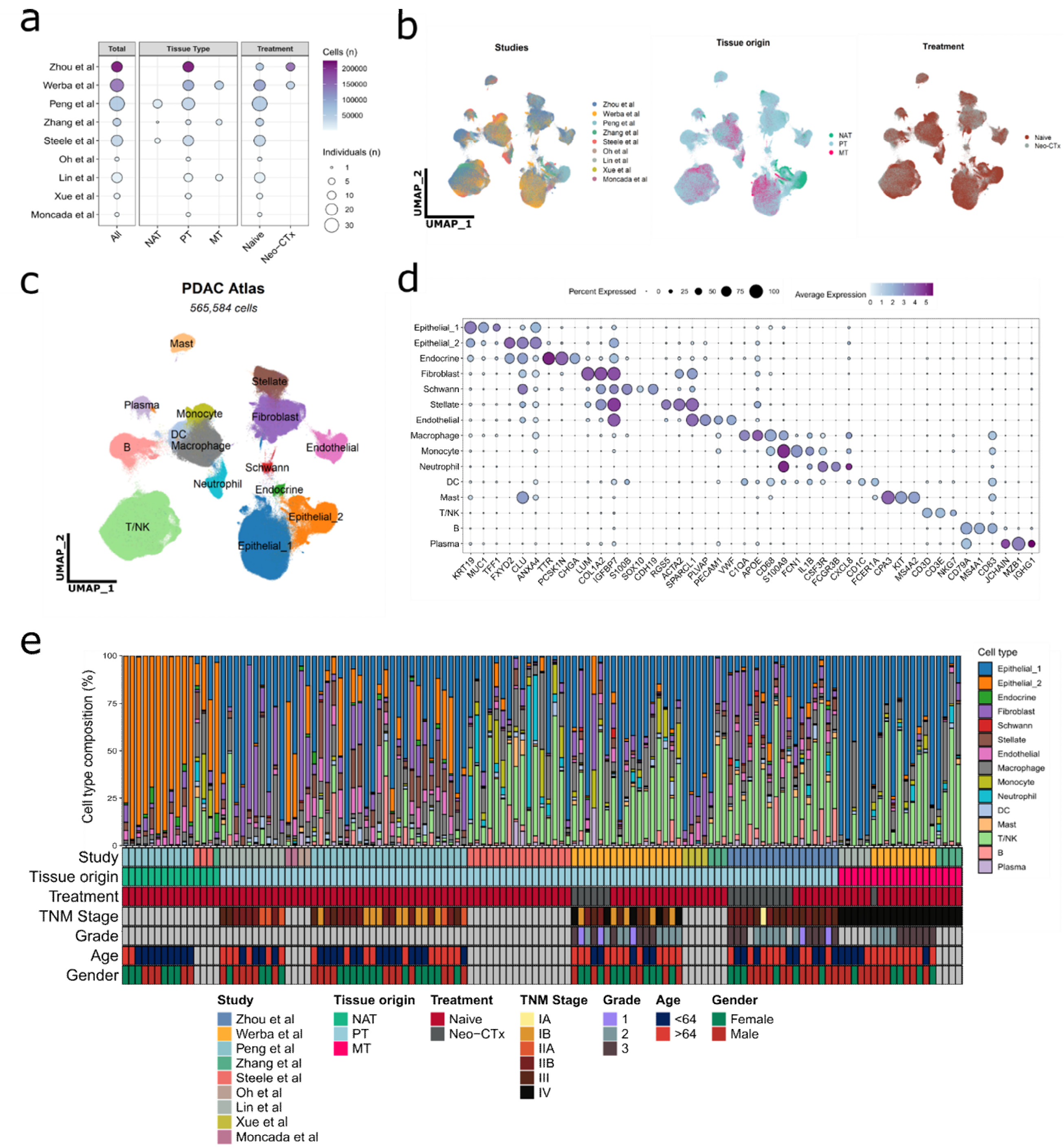


Figure 3.3 Broad cellular landscape of PDAC atlas. (a) Summary statistics of nine PDAC datasets separated by tissue types and treatment conditions. Dot plot size represents the number of individuals and colour represents the number of total cells. **(b)** UMAP embeddings of PDAC atlas coloured by study (left), tissue origin (middle), and neo-CTx treatment (right). **(c)** UMAP embedding of PDAC atlas coloured by the 15 broad cell types. **(d)** Dot plots showing expression levels and frequencies of representative canonical genes. **(e)** Cell type composition across all samples with annotated clinical features.

Next, cell type abundance dynamics across tissue types and treatment conditions were investigated. Epithelial cluster 1 (Epithelial_c1) cells represented mainly malignant cells based on PDAC-specific markers (*KRT19, MUC1,* and *TFF1)* (Fig. 3.3d), and were significantly enriched in both PT and MT tissues compared to NAT tissue (Fig. 3.4a-b). Additionally, epithelial_c1 cells were significantly enriched in MT compared to PT tissues, possibly due to differences in tissue preparation as MT tissues were obtained using targeted biopsies enriched for malignant cells. In contrast, epithelial cluster 2 (Epithelial_c2) represented normal ductal/acinar cells and were significantly depleted in PT and MT tissues compared to NAT (Fig. 3.4a-b).

PT tissues also displayed enrichment of fibroblasts, macrophages, dendritic cells and T/NK cells. Notably, T/NK cells were the most abundant immune population in both PT and MT tissues, aligning with previous reports[137].

Moreover, post neo-CTx samples showed a significant increase in T/NK and mast cells and a reduction of myeloid cells, including neutrophils, suggesting adaptive immune activation (Fig. 3.4c-d). Stromal cells were also significantly enriched, including fibroblasts and Schwann cells, highlighting stromal cell remodelling following neo-CTx (Fig. 3.4c-d).

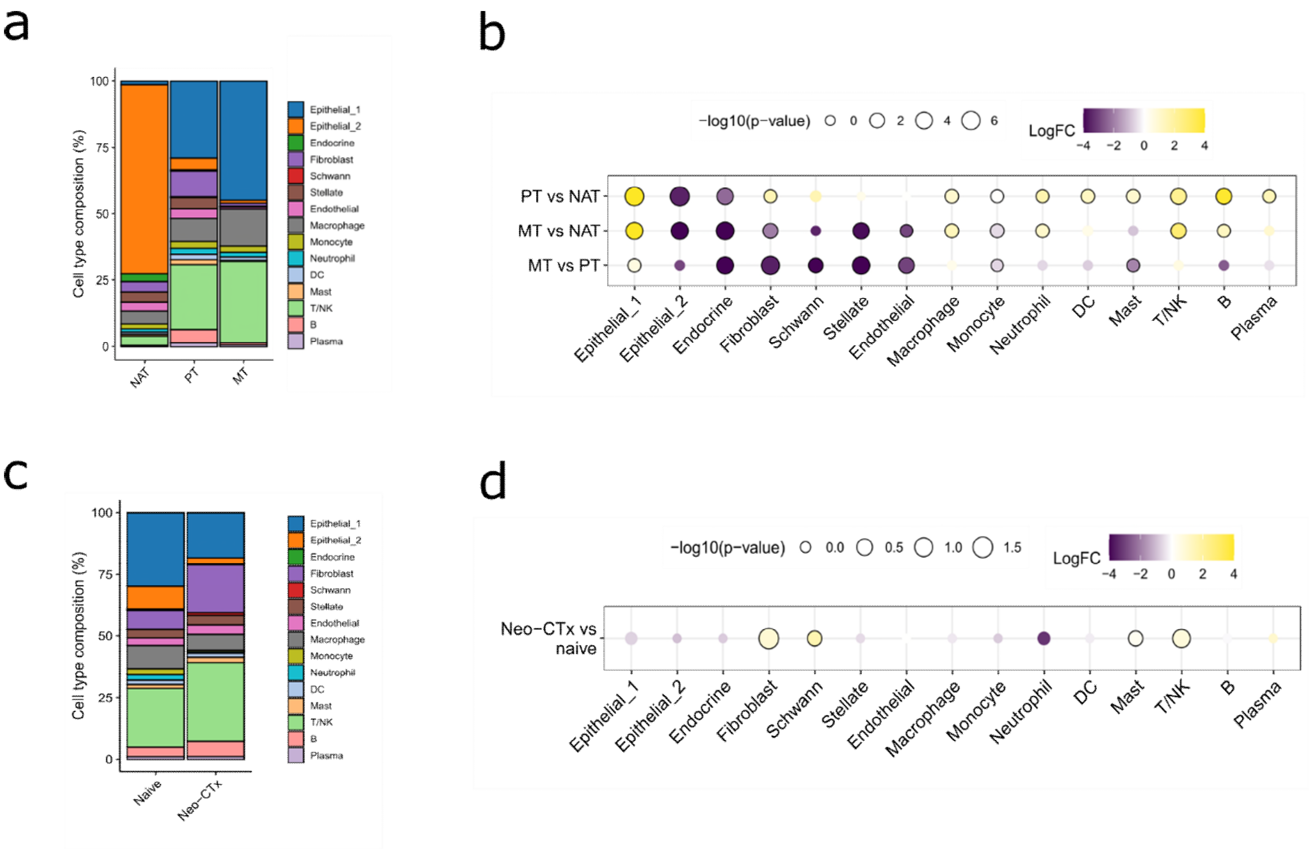


Figure 3.4 Broad cell type composition analysis. (a) Cell type composition across NAT, PT, and MT tissue types. **(b)** Dot plots showing the changes of cell type abundance in PT and MT tissues compared to NAT and MT compared to PT. Outlined dots represent significant differences ($P < 0.05$). Statistical significance was calculated using a two-sided Wilcoxon rank-sum test. **(c)** Cell type composition across treatment-naïve and neo-CTx treated samples. 4.5 TME cell state characterization. **(d)** Dot plots showing the changes in cell type abundance in neo-CTx against treatment-naïve samples. Outlined dots represent significant differences ($P < 0.05$). Statistical significance was calculated using a two-sided Wilcoxon rank-sum test.

3.5 Identification and characterization of TME cell states

We further subclustered the immune and stromal cells revealing 53 cell states, which include, 7 CD4+ T, 10 CD8+T and NK, 6 B and plasma, 16 myeloid, and 14 stromal cell states (Fig. 3.5a). Unsupervised hierarchical clustering was performed to inspect the transcriptional similarity of these cell states grouping them into 8 major branches, consisting of proliferating lymphoid, T/NK, endothelial, stromal, B/plasma, mast, neutrophils, and myeloid cells (Fig. 3.5b).

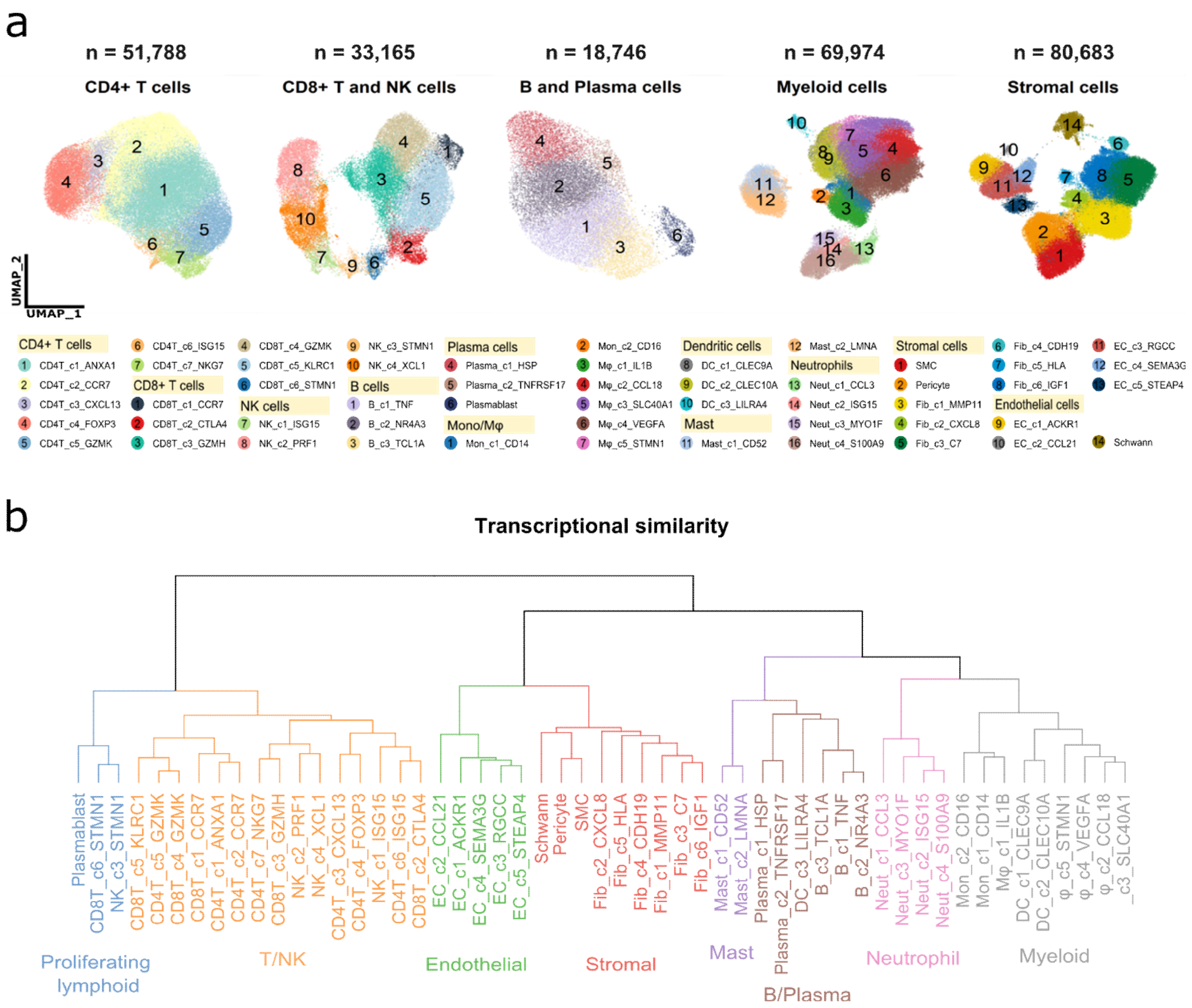


Figure 3.5 TME cell state landscape. (a) UMAP embeddings of CD4+ T, CD8+T/NK, B/plasma, myeloid, and stromal cell states. **(b)** Unsupervised hierarchical clustering of defined TME cell states coloured into 8 major branches.

These cell states displayed unique gene expression profiles which are described in the following subsections below.

3.5.1 Lymphoid cell state landscape

The 7 CD4+ T cell states included: T helper 1 (Th1)-like (CD4T_c1_ANXA1), naïve (CD4T_c2_CCR7), follicular helper (Tfh) (CD4T_c3_CXCL13), Treg (CD4T_c4_FOXP3), activated (CD4T_c5_GZMK), interferon (IFN) response (CD4T_c6_ISG15), and cytotoxic (CD4T_c7_NKG7) (Fig. 3.6a and Supplementary Table 2).

Moreover, 6 CD8+ T cell states included: naive (CD8T_c1_CCR7), dysfunctional (CD8T_c2_CTLA4), cytotoxic (CD8T_c3_GZMH) transitional effector (CD8T_c4_GZMK), resident memory (CD4T_c5_KLRC1), and proliferating (CD8T_c6_STMN1) states (Fig. 3.6b and Supplementary Table 3).

Furthermore, the four NK cell states included: IFN response (NK_c1_ISG15), cytotoxic (NK_c2_PRF1), proliferating (NK_c3_STMN1), and chemokine-secreting (NK_c4_XCL1). (Fig. 3.6b and Supplementary Table 4). The NK_c1_ISG15 subcluster displayed the highest expression of *HAVCR2 (*TIM-3), and *TIGIT* checkpoint receptors among NK cells, reflecting a dysfunctional phenotype similar to CD8T_c2_CTLA4. Both NK_c1_ISG15 and NK_c4_XCL1 represented sub-clusters of the tissue-resident NK (trNK) population, which we previously characterized in PDAC[138].

Lastly, the six B and plasma cell states included: transitional activated (B_c1_TNF), activated (B_c3_NR4A3), naïve (B_c3_TCL1A), stressed plasma (Plasma_c1_HSP), mature plasma (Plasma_c2_TNFRSF17), and proliferating plasmablast (Fig. 3.6c and Supplementary Table 5).

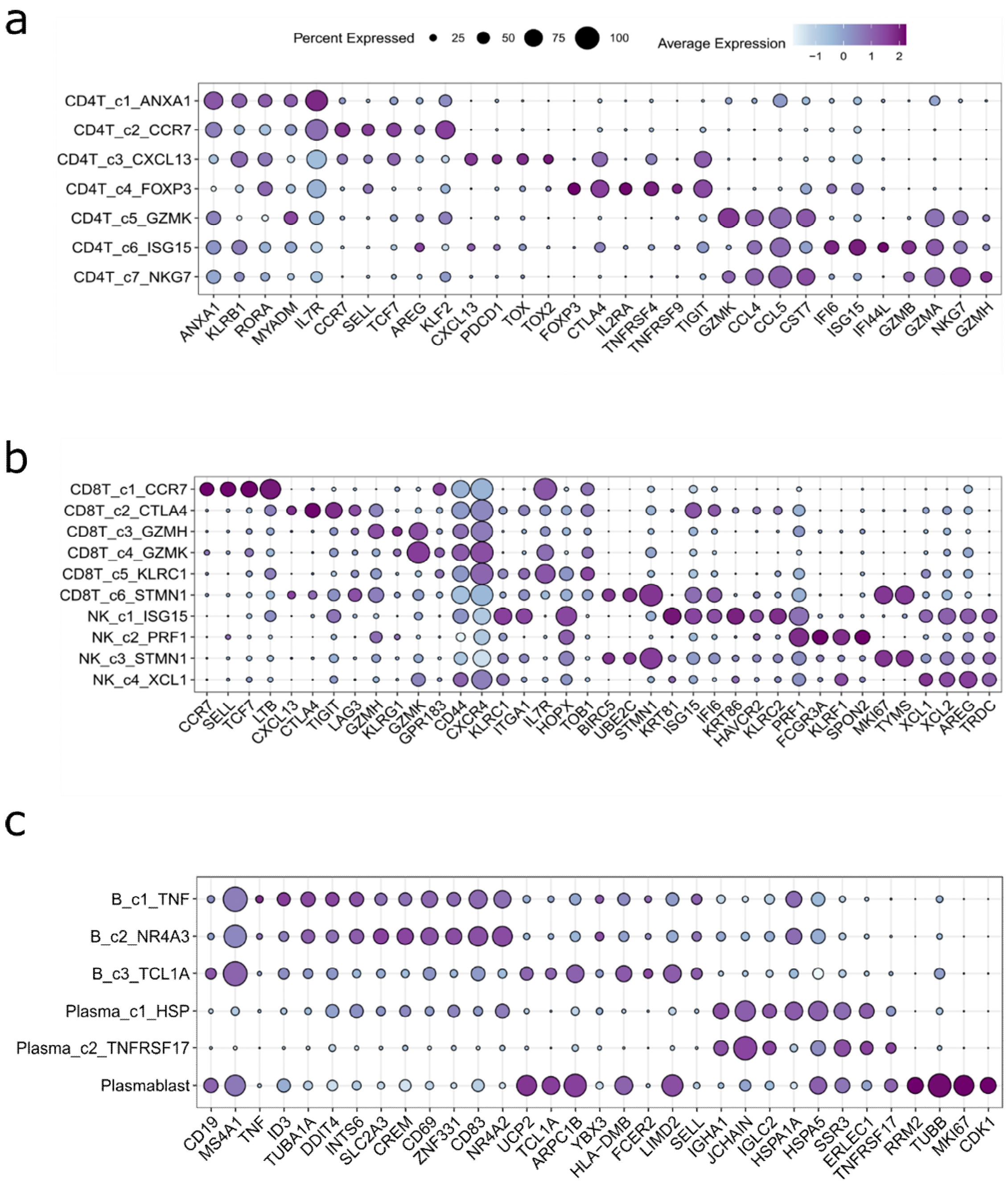


Figure 3.6 Characterization of lymphoid lineage cell states. (a) Dot plot showing expression levels and frequencies of representative signature genes for CD4+ T cell states. **(b)** Dot plot showing expression levels and frequencies of representative signature genes for CD8+ T and NK cell states. **(c)** Dot plot showing expression levels and frequencies of representative signature genes for B and plasma cell states.

3.5.2 Myeloid cell state landscape

The two monocyte subclusters included: classical (Mon_c1_CD14), and non-classical (Mon_c2_CD16*)*, and 5 macrophage subclusters: inflammatory (Mϕ_c1_IL1B), lipid-associated (Mϕ_c2_CCL18), complement-enriched (Mϕ_c3_SLC40A1), hypoxic (Mϕ_c4_VEGFA), and proliferating (Mϕ_c5_STMN1*)* (Fig. 3.7a and Supplementary Table 6). The Mϕ_c1_IL1B displayed the highest 'M1' macrophage signature score while the lipid-associated Mϕ_c2_CCL18 exhibited the highest 'M2' signature and 'Checkpoint ligands' scores (Fig. 3.7b). Specifically, the Mϕ_c2_CCL18 exhibited the highest expression and frequency of *LGALS9* (Galectin-9), *SIRPA,* and *LAIR1* checkpoint molecules, highlighting its immunosuppressive phenotype (Fig. 3.7c). PD-L1 (encoded by *CD274*) expression was low compared to other checkpoint ligands in all macrophage populations, and could explain the poor clinical responses to anti-PD-L1 ICB in PDAC[139].

Moreover, we identified three dendritic cell states: conventional DC 1-like (cDC1) (DC_c1_CLEC9A), cDC2-like (DC_c2_CLEC10A), and plasmacytoid (pDC) (DC_c3_LILRA4) (Supplementary Table 7). Furthermore, we identified two mast cell populations, Mast_c1_CD52 expressing high *CD52*, and Mast_c2_LMNA expressing high *LMNA,* and *RHEX*. Notably, mast cells specifically expressed *CSF1* which binds to its cognate receptor *CSF1R*, involved with TAM infiltration and immunosuppression[140] (Supplementary Table 8). Lastly, we identified four neutrophil subclusters: hypoxic (Neut_c1_CCL3), IFN response (Neut_c2_ISG15), mitochondrial-gene enriched (Neut_c3_MYO1F), and neutrophil-extracellular trap (NET)-associated (Neut_c4_S100A9) phenotypes (Fig.3.7a, d and Supplementary Table 8). The Neut_c2_ISG15 displayed the highest enrichment of the tumour-associated neutrophil (TAN) gene signature.

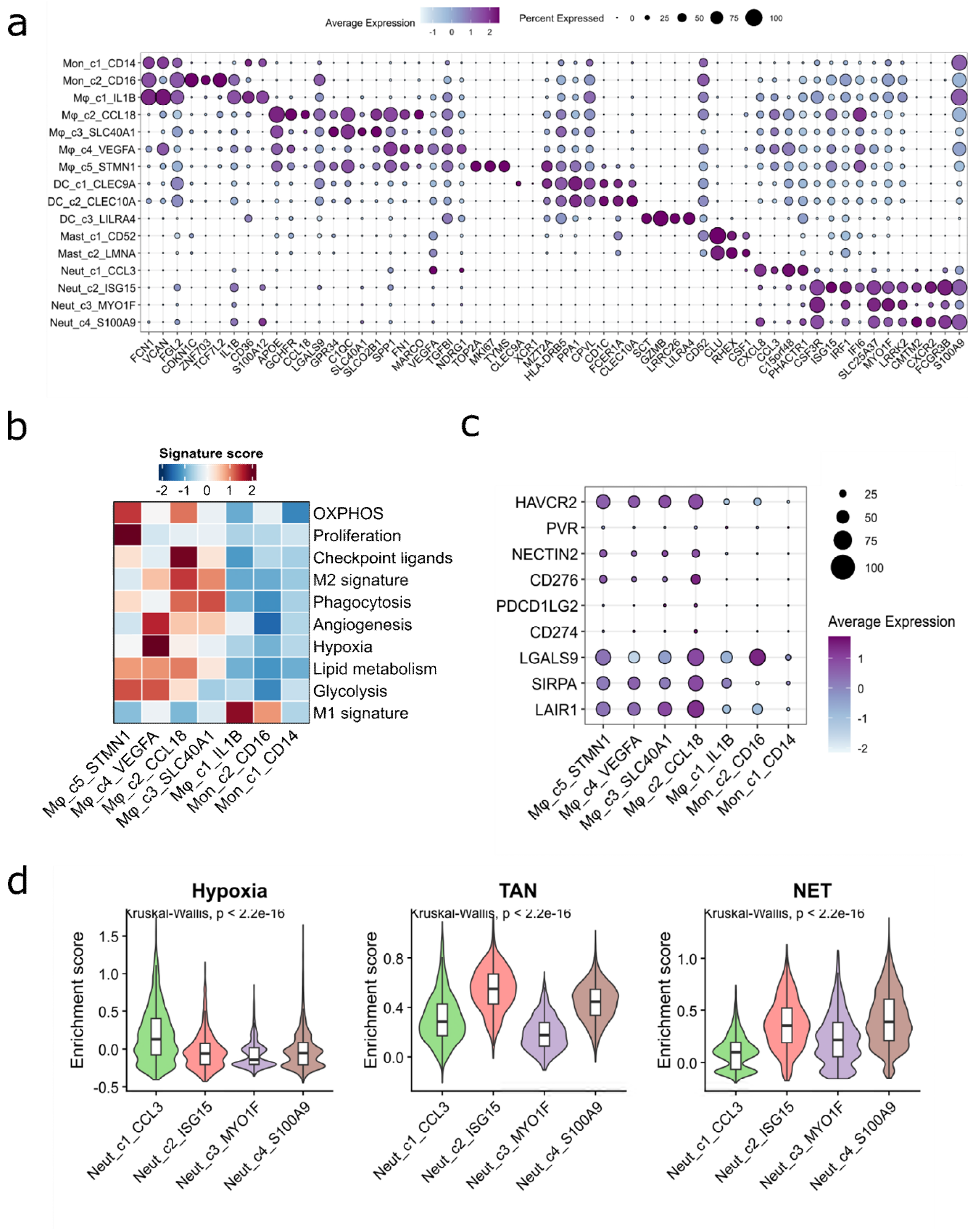


Figure 3.7 Characterization of myeloid cell states. (a) Dot plot showing expression levels and frequencies of representative signature genes for myeloid cell states. **(b)** Heatmap showing the scaled

enrichment scores of macrophage-specific gene signatures. **(c)** Dot plot showing the expression levels and frequencies of selected checkpoint ligands across monocyte and macrophage subclusters. **(d)** Violin plots showing the enrichment scores of selected gene signatures across neutrophil subclusters. Statistical significance was determined using the non-parametric Kruskal Wallis test. OXPHOS: Oxidated Phosphorylation; TAN: Tumour-Associated Neutrophils; NET: Neutrophil Extracellular Traps.

3.5.3 Stromal cell state landscape

The 8 stromal cell subclusters included: smooth muscle cells (SMCs), pericytes, ECM-related (Fib_c1_MMP1), hypoxic (Fib_c2_CXCL8), complement-enriched (Fib_c3_C7), nerve-like (Fib_c4_CDH19), antigen-presenting (Fib_c5_HLA), and a second complement-enriched fibroblast subcluster (Fib_c6_IGF1) also expressing high *IGF1, PDGFRA,* and *TNXB* (Fig. 3.8a and Supplementary Table 9).

Notably, certain fibroblast cell states strongly overlapped with previous CAF classifications (Fig. 3.8b). For instance, the Fib_c1_MMP11 overlapped strongly with the ‘myCAF’ and ‘myofibroblastic’ signatures, and was significantly enriched with the ‘Collagen-fibril organization’ signature (Fig. 3.8c). The Fib_c2_CXCL8 displayed poor overlap with the three ‘myCAF’, ‘iCAF’, and ‘apCAF’ signatures, however it was enriched with the ‘Immunomodulatory’ signature. Additionally, this subcluster displayed elevated levels of *SERPINE1* (encoding PAI-1) which is an inhibitor of fibrinolysis, and displayed enrichment of the ‘Negative regulation of fibrinolysis’ signature (Fig. 3.8a-c). Fib_c3_C7, Fib_c4_CDH19, and Fib_c5_IGF1 overlapped with the ‘Neurotropic’ gene signature, and shared enriched pathways, including ‘Synaptic transmission’, and ‘Complement activation’ (Fig. 3.8c). Moreover, the Fib_c5_IGF1 displayed the strongest enrichment of the ‘iCAF’ phenotype. The Fib_c5_HLA strongly overlapped with the 'apCAF' gene signature, and was enriched for 'Antigen processing and presentation of peptide antigen' and 'Natural killer cell chemotaxis' gene signatures, reflecting immunomodulatory properties (Fig. 3.8c).

Lastly, the four endothelial cell subpopulations included: venular (EC_c1_ACKR1), lymphatic (EC_c2_CCL21), capillary (EC_c3_RGCC), arterial (EC_c4_SEMA3G), and an *RGS5+* stress response population (EC_c5_STEAP4) (Fig. 3.8a and Supplementary Table 10).

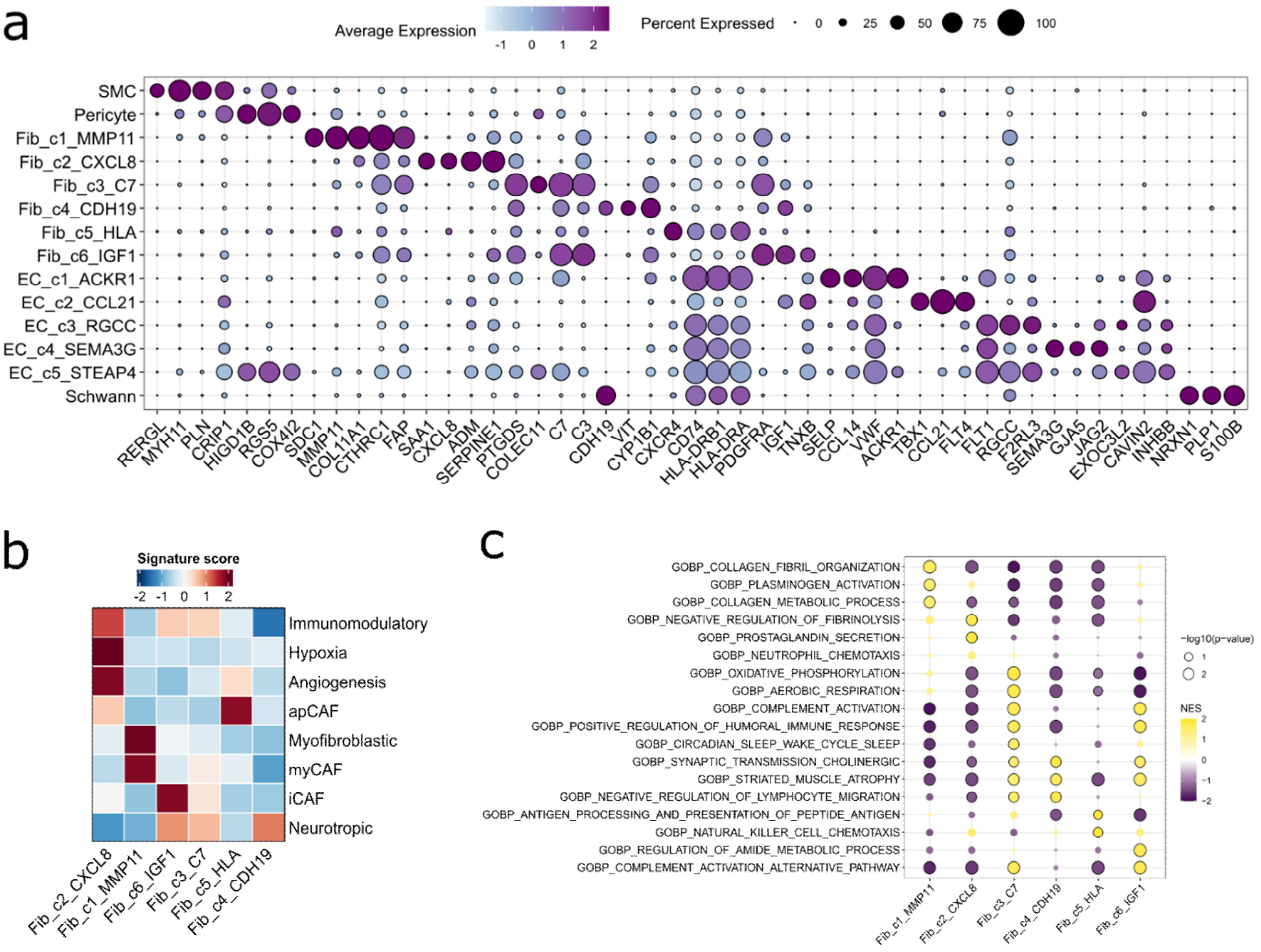


Figure 3.8 Characterization of stromal cell states. (a) Dot plot showing expression levels and frequencies of representative signature genes for stromal cell states. **(b)** Heatmap showing the scaled enrichment scores of fibroblast-specific gene signatures. **(c)** Dot plot showing selected gene sets enriched in the fibroblast cell states following gene set enrichment analysis. Colour of dots represent the

normalized enrichment score (NES) and the size represents the -log10 (p-value). Outlined dots represent significant differences ($P < 0.05$).

3.6 Association of TME cell states with clinical features

3.6.1 Association with disease progression

Cell state abundance dynamics across NAT, PT, and MT tissue were investigated. Cell states of major cell types displayed diverse tissue preferences. For CD4+ T cells, both CD4T_c4_FOXP3 and CD4T_c6_ISG15 were significantly enriched in PT and MT tissues while naïve CD4T_c2_CCR7 cells were mainly depleted, though not significant (Fig. 3.9a). Notably, both helper CD4T_c1_ANXA1 and CD4T_c5_GZMK were significantly enriched in MT vs PT tissues. The majority of CD8+ T and NK cells were significantly enriched in PT and MT tissues compared to NAT. In contrast, the resident memory population (CD8T_c5_KLRC1) was enriched in NAT and depleted in both PT and MT tissue, consistent with previous reports of high $CD8^+$ T resident memory (Trm) cell abundance in non-malignant tissues[141 142].

Moreover, myeloid cells displayed variable compositions in tissue types. The Mϕ_c2_CCL18 population was significantly enriched in PT and MT against NAT as well as MT against PT, consistent with previous findings[143] (Fig. 3.9a). In contrast, both DC_c1_CLEC9A and DC_c2_CLEC10A populations were depleted in both PT and MT tissues, reflecting loss of effective antigen presentation during disease progression, in line with previous mouse model studies[144].

Furthermore, both SMCs and Fib_c6_IGF1 were depleted in PT and MT tissues, while pericytes and myCAF (Fib_c1_MMP11) were enriched in both PT and MT tissues. Most endothelial cells were enriched in MT compared to PT tissues, highlighting the sparse vasculature of PDAC primary tumours compared to metastatic sites[145].

We next utilised MultiNicheNet[111], a novel cell-cell communication framework for comparing L-R activities across conditions. Using this tool, the top 50 L-R activities between PT and MT tissues were investigated. Results show that fibroblasts were the main senders in PT tissues, while macrophages and NK cells were the main senders in MT tissues (Fig. 3.9a-b). For instance, complement C3-C3AR1 from fibroblasts to macrophages was elevated in PT tissues, and this signalling has recently been shown to drive tumour progression and ICB response in gastric cancer[146]. Interestingly, PT tumours were enriched with CSF1-CSF1R originating from Mast and NK cells towards DCs but not macrophages (Fig. 3.9b).

In MT tissues, there were enriched chemokines *CCL3/4/8* originating from macrophages, NK cells, and DCs mainly towards macrophages (Fig. 3.9c). Notably, the AREG-MMP9 signalling was upregulated in MT tissue, originating from both NK and DCs towards macrophages, and MMP9 activation has been shown to downregulate phagocytosis-induced cell death (PICD)[147]. Additionally, elevated CXCL12-CXCR4 originating fibroblasts towards neutrophils was identified, suggesting a role in neutrophil recruitment and immunosuppression at metastatic sites[148]. Taken together, these findings highlight extensive remodelling of TME cell states and intercellular communication networks throughout PDAC disease progression.

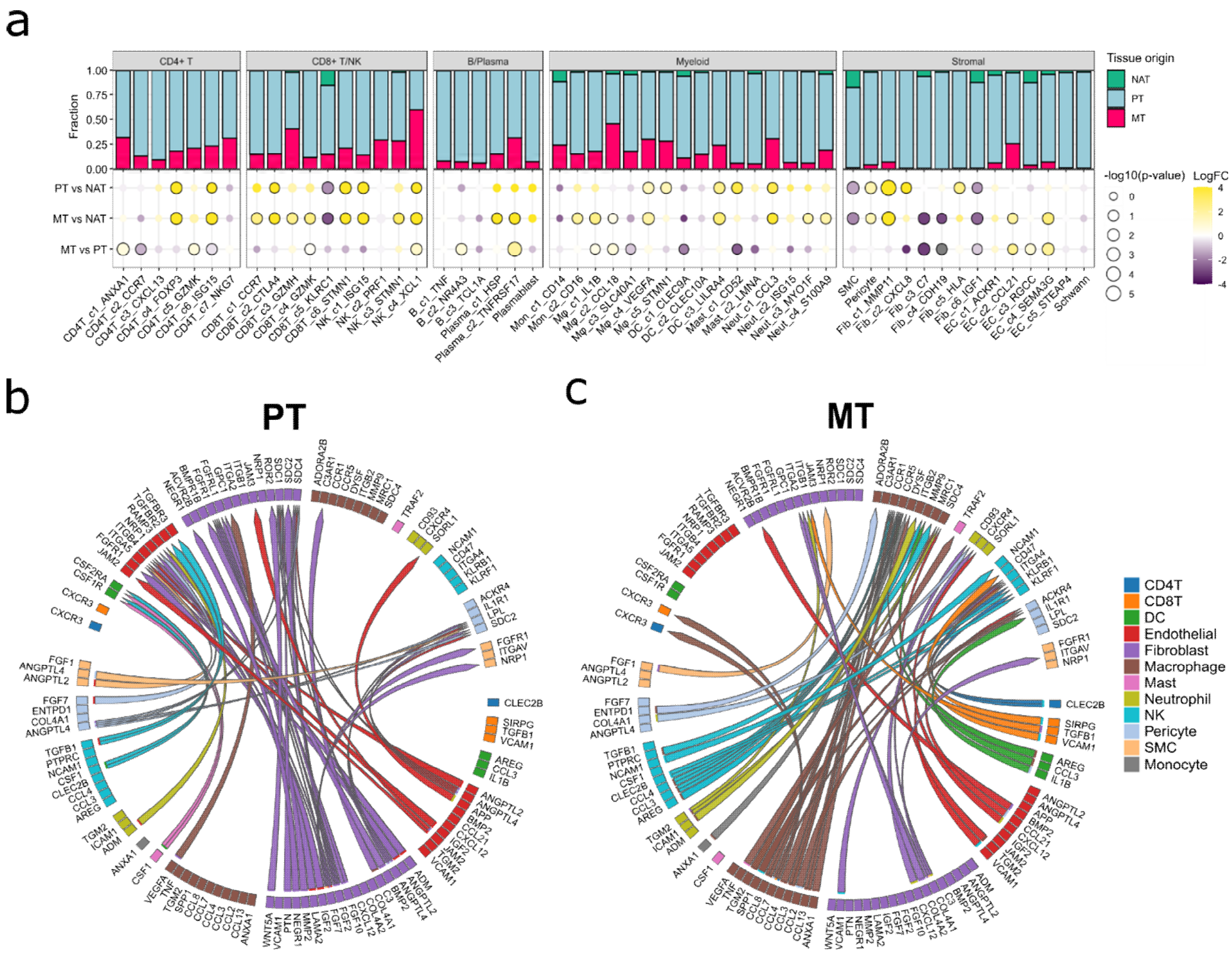


Figure 3.9 Association of TME cell landscape and disease progression. (a) Dot plot showing the changes of cell state abundance in PT and MT tissues compared to NAT and MT compared to PT. Outlined dots represent significant differences ($P < 0.05$). Statistical significance was calculated using a two-sided Wilcoxon rank-sum test. Bar plots above show the tissue origin composition for each cell state **(b)** Cell-cell interaction plot showing ligand-receptor pairs enriched in PT tissue. **(c)** Cell-cell interaction plot showing ligand-receptor pairs enriched in MT tissue.

3.6.2 Association with neo-CTx

Chemotherapy has been shown to shape the tumour microenvironment of pancreatic cancer altering communication networks and influencing prognosis[149]. In our analysis, we compared the abundance and intercellular communication of cell types between patient samples that received either neo-adjuvant FOLFIRINOX (n = 12), Gemcitabine/nab-paclitaxel (n = 4), or chemoradiation (n = 2) and those that were treatment-naïve (n = 108). neo-CTx significantly increased the abundance of activated CD4+ T cells (CD4T_c1_ANXA1, CD4T_c3_CXCL13, CD4T_c5_GZMK), and CD8+ T cells (CD8T_c4_GZMK, CD8T_c5_KLRC1), along with chemokine secreting NK_c4_XCL1 (Fig. 3.10a). In contrast, there was a reduction in the immunosuppressive Mϕ_c2_CCL18 population. This is consistent with the reduction of checkpoint L-R interactions towards CD8+ T cells following chemotherapy, that may promote resistance to immunotherapy[133].

Moreover, there was a significant increase in complement-secreting (Fib_c3_C7, Fib_c6_IGF1) and neurotropic (Fib_c4_CDH19) fibroblasts, Schwann cells, and mast cells. Elevated TME complement following neo-CTx has recently been reported in PDAC and associated with improved overall survival[150].

Communication analysis shows enriched SERPINE1-PLAUR and PLAU-PLAUR signalling from fibroblasts towards neutrophils in the treatment-naïve group, pathways previously shown to promote neutrophil infiltration[151] (Fig. 3.10b). Moreover, Oncostatin M signalling (OSM-OSMR*)* originating from neutrophils towards Schwann cells was also elevated in the naïve group. This myeloid-derived signal has been reported to reprogramme stromal cells toward a tumorigenic phenotype promoting cancer progression[152]. Collectively, these findings suggest that multiple immunosuppressive and tumour-promoting interactions within the TME are downregulated following neo-CTx.

In contrast, the neo-CTx group displayed enriched growth factors from Schwann cells towards fibroblasts and SMCs, including PDGFB-PDGFRA, HBEGF-EGFR, and FGF2-FGR1 (Fig 3.10c). Notably, an increase in IFNG-IFNGR1 signalling from NK, CD4+, and CD8+ T cells towards neutrophils was identified. Interferon gamma has been shown to increase NET formation leading to tumour-killing activity in colorectal cancer[153].

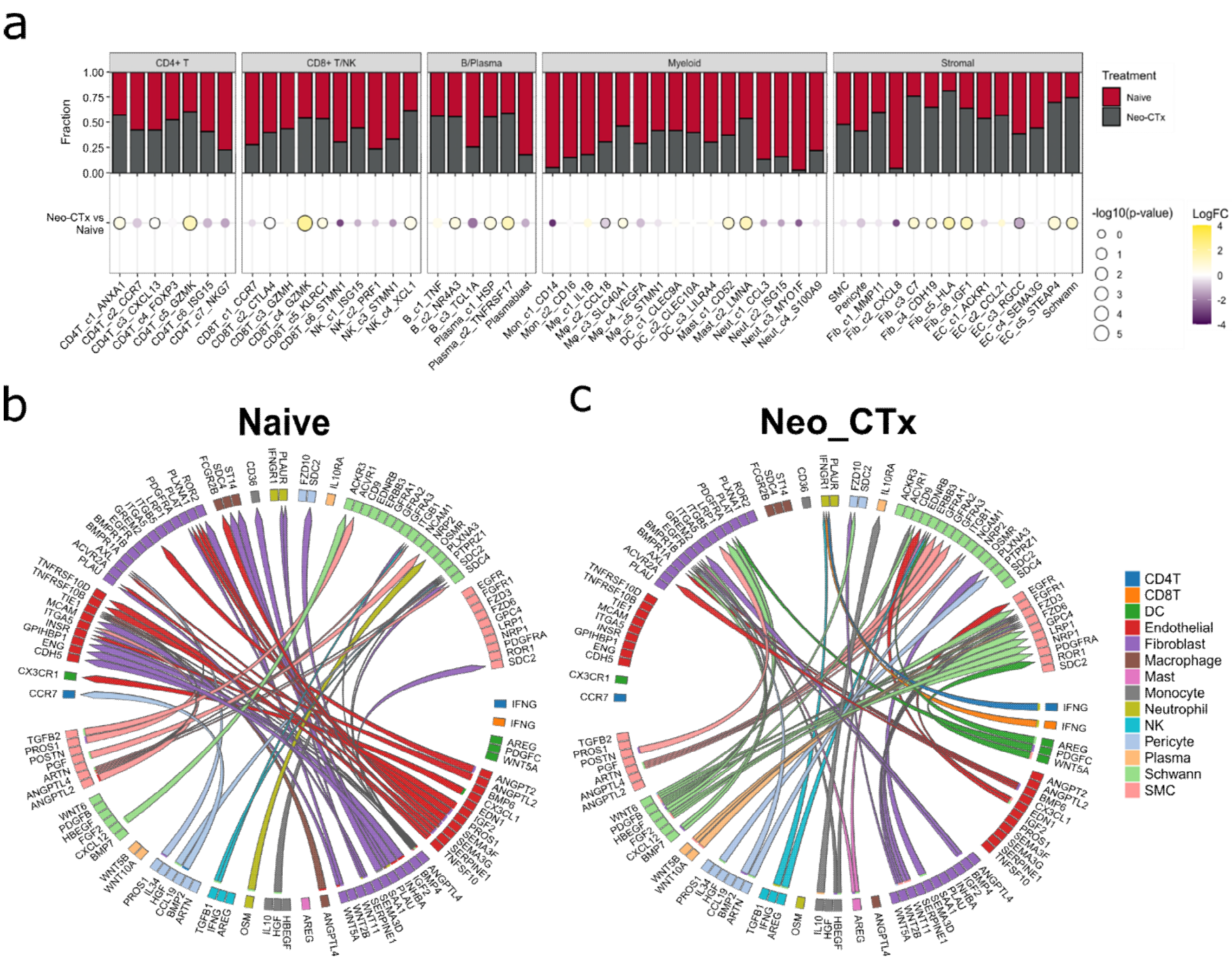


Figure 3.10 Association of TME cell landscape and neo-adjuvant therapy. (a) Dot plot showing the changes of cell state abundance in neo-CTx treated samples compared treatment-naive. Outlined dots represent significant differences (*P* < 0.05). Statistical significance was calculated using a two-sided Wilcoxon rank-sum test. Bar plots above show the treatment composition for each cell state **(b)** Cell-cell interaction plot showing ligand-receptor pairs enriched in treatment-naive samples. **(c)** Cell-cell interaction plot showing ligand-receptor pairs enriched in neo-CTx treated samples.

3.6.3 Association with clinical features

Given the scale of the single-cell atlas and available clinical information, we investigated the association of TME cell states with additional clinical features, including age (n = 52, >64 age = 27, <64 age = 25), gender (n = 52, female = 26, male = 26), and histological grade (n = 18, grade 3 = 8, grades 1+2 = 10). When comparing with age, patients above the median age (64 years old) displayed a reduction of resident memory and proliferating CD8+ T cells (CD8T_c5_KLRC1, CD8T_c6_STMN1) as well as reduced cDC1 (DC_c1_CLEC9A) validating the reduced antitumor immunity previously reported with aging[154] (Fig. 3.11a). Notably, we observed a significant reduction of apCAFs (Fib_c5_HLA), a finding not previously reported in PDAC that warrants further investigation.

Moreover, comparison of cell state abundance by gender revealed higher levels of Tfh (CD4T_c3_CXCL13) cells in females (Fig. 3.11b). Interestingly, elevated serum estradiol (E2) levels have recently been associated with increased Tfh frequencies[155], suggesting a potential role of oestrogen signalling in modulating T cell function. There was a slight enrichment of naïve B cells and plasmablasts in females, though not significant. Additionally, results show increase in Fib_c6_IGF1 and SMC populations in females compared to males (Fig. 3.11b). This is consistent with the recent association of oestrogen signalling and stromal cell abundance, where β-oestradiol resulted in a shift towards the iCAF phenotype[156].

Lastly, comparison of cell state abundance with histological tumour grade did not reveal any significant correlations as number of grade groups were limited (G3:8 and G1+G2:10). However, grade 3 tumours showed reduced M1-like macrophage (Mϕ_c1_IL1B, $p = 0.07$), and monocyte (Mon_c1_CD14, $p = 0.05$) abundance, which could indicate phenotype-dependent cancer cell remodelling of immune cells.

Figure 3.11 Association of TME cell state abundance and clinical features. (a) Dot plot showing the changes of cell state abundance between >64-year-old and <64-year-old patients. **(b)** Dot plot showing the changes of cell state abundance between females and males. **(c)** Dot plot showing the changes of cell state abundance between higher grade (G3) and lower histological (G1+2) tumour grades. Outlined dots represent significant differences ($P < 0.05$). Statistical significance was calculated using a two-sided Wilcoxon rank-sum test.

3.7 Conclusions and Discussion

In this chapter we constructed an integrated reference atlas of PDAC encompassing a variety of disease states, and treatment conditions. Importantly, we utilized the integration tool scANVI, which was the top-performer when compared with other tools for its accuracy in bio-conservation and batch correction.

Given the scale of PDAC single-cell atlas, we identified and characterized a diverse range of cell states across major cell types, each defined by distinct gene expression profiles. While Elyada et al. previously described three CAF populations (myCAF, iCAF, and apCAF)[56], we identified further sub-clusters. Notably, the Fib_c3_C7 and Fib_c6_IGF1 were enriched for the iCAF signature, with the latter predominantly present in normal/adjacent pancreatic tissue (NAT). Moreover, the hypoxic Fib_c2_CXCL8 population, enriched in neutrophil-recruiting chemokines, strongly aligned with the "immunomodulatory" program[61], alongside the two iCAF-like clusters (Fib_c3_C7 and Fib_c6_IGF1). Interestingly, we revealed a previously uncharacterized *CDH19*$^+$ fibroblast population (Fib_c4_CDH19), that represented a subset of both the iCAF and neurotropic cell states. CDH19 is a known marker for Schwann cells[157], however this population formed a separate cluster from Schwann cells. However, they may represent Schwann cell precursors (SCPs) originating from fibroblast populations[158], given that they are located between both cell populations in the integrated UMAP embeddings (Fig. 3.5a).

Immunosuppression is a hallmark of PDAC, and elucidating the intercellular communications might help overcome resistance to immunotherapy[159]. Results show that the metabolically-active Mϕ_c2_CCL18 population, characterized by the highest expression of checkpoint ligands, was enriched in metastatic tissue, indicating elevated immune suppression in PDAC liver metastases. This aligns with the enrichment of *MRC1+ CCL18+* macrophages in colorectal cancer liver metastases[160]. Macrophage-specific targeting with an anti-CSF-1R antibody in combination with immune checkpoint blockade has shown limited efficacy in advanced CRC and PDAC[161]. Our analysis identified elevated CSF1-CSF1R signalling directed towards DCs in

PT tissues, which may explain why the CSF-1R inhibitor pexidartinib, led to DC depletion and poor antitumour clinical activity in advanced CRC and PDAC patients[161]. In contrast, macrophage-targeted CSF1-CSF1R signalling was not detected in MT tissues, suggesting alternate targets may be more effective in metastatic PDAC. Notably, chemokines *CCL3*, *CCL4*, and *CCL8* were enriched in MT tissues. Thus, targeting their shared receptor CCR5 with the inhibitor maraviroc may enhance therapeutic outcomes by reducing TAMs, MDSC, and Treg abundance[162 163]. Taken together, these findings highlight substantial TME remodelling and communications between PT and MT tissues with clinical and therapeutic significance.

Neo-adjuvant CTx has been shown to modulate the TME, thus affecting immunotherapeutic responses[164]. For instance, Werba et al shown a higher *TIGIT* expression in CD8+ T cells in treatment-naïve samples, thus proposing that TIGIT blockade may be more effective as first-line treatment or in combination with chemotherapy[133]. Our findings show a reduction of the immunosuppressive Mϕ_c2_CCL18 population in neo-CTx treated samples with an increase in anti-tumour phenotypes such as *GZMK*+ CD8+ T and CD4+ T cells and *XCL1*+ NK cells. This is line with Neo-CTx responsive tumours in colorectal cancer[160]. However, in tumours non-responsive to Neo-CTx, there was enrichment of *SPP1*+ macrophages and *MRC1*+ *CCL18*+ macrophages and downregulation of cytotoxic immune cells[160]. This opposite effect of immune reprogramming may be driven by the chemotherapeutic remodelling of cancer cells and this chemo-modulation effect has also been reported in an *ex vivo* culture platform[165]. Specifically, in CRC samples non-responsive to neo-CTx, there was a reported increase in fatty acid (FA) metabolism in cancer cells[160]. Thus, a possible mechanism is that increase in CTx-resistant cancer cells could metabolically reprogram macrophages to an M2 CCL18+ phenotype thus driving immunosuppression and cancer cell progression[166]. Sohal and others have reported differential immune modulation between neoadjuvant FOLFIRINOX or GemCap chemotherapy[164], including a more pronounced reduction of proliferating cells following FOLFIRINOX treatment. However, our analysis compared all neo-CTx regimens collectively against treatment-naïve samples. Future work dissecting regimen-specific effects on TME cell state modulation in PDAC is therefore warranted.

In summary, our findings highlight pronounced immunosuppressive mechanisms enriched in PDAC metastases, as well as restoration of anti-tumour immune activity following neo-CTx in treatment-responsive patients. Together, these observations support the rationale for combination therapies, including metabolic interventions, such as PIKfyve-mediated disruption of lipid metabolism[167], or immune checkpoint blockade in treatment-naive patients exhibiting elevated M2 CCL18+ macrophage levels.

However, our analysis did have some limitations. For instance, only the top 50 ranked L-R interactions were evaluated for both tissue type and neo-CTx analyses, potentially overlooking other relevant interactions. Additionally, due to the limited number of neo-CTx samples, we were unable to stratify patients based on chemotherapy response. Future integration of additional neo-CTx treated samples with follow-up data into the reference atlas, using approaches such as the scArches method[168], could enable rapid and comprehensive annotation of new datasets, and facilitate more efficient downstream analyses.

Chapter 4

Discovery of PDAC multi-cellular ecosystems

4.1 Introduction and Aims

The cancer TME consists of multi-cellular ecosystems with intricate signalling networks that drive tumour progression, metastasis, and treatment responses[169 126]. Previous studies have revealed broad TME classifications, including T cell inflamed ("hot") and immune excluded/fibrotic ("cold") tumours[169]. While these classifications can inform clinical outcomes, such as responses to ICB, they oversimplify the diversity of cellular states within the TME. In recent years, single-cell and spatial transcriptomics have emerged as powerful approaches for mapping high-resolution cellular communities[126], however, most studies are limited by small cohort sizes preventing robust analyses. An alternative strategy is estimating cell type or cell state abundances by deconvoluting bulk RNA-seq datasets. For example, the recent EcoTyper machine learning framework has been applied across cancer types to identify multicellular communities strongly associated with overall survival and responses to ICB[170]. However, the accuracy of cellular deconvolution in bulk specimens is largely influenced by the choice of deconvolution method and cell type definitions[171].

Given the large sample size in our PDAC single cell atlas and the diverse immune and stromal cell states, we aim to identify and characterize robust CEs consisting of co-occurring cell states based on absolute frequencies. The specific aims of this chapter include:

- Identify robust PDAC CEs using our bimodal analysis approach (Methods 2.6).

- Characterize CEs based on clinical features and gene signature profiles.
- Investigate the spatial characteristics of CEs.
- Investigate the multicellular communication of CEs.

4.2 Identification of five robust CEs

To identify multi-cellular ecosystems, we investigated covariance in cell state frequencies across samples using two parallel modules (Fig. 2.2 and Methods 2.6). For the first module, we applied consensus clustering to the normalized cell state frequency matrix using the PAM algorithm. Based on the highest ASW, five consensus clusters were identified as optimal with compact Jaccard similarity index (Fig. 4.1a-b). Principal component analysis (PCA) suggested clear separation among the five clusters in two-dimensional spatial coordinates with minimal overlap (Fig. 4.1c). Next, cell states enriched in each consensus cluster were prioritized based on a statistical ranking metric (logFC x -log10(p-value)) derived from a two-sided Wilcoxon test (Fig. 4.1.d).

For the second module, we computed pairwise correlations between all cell state pairs, retaining only positive correlations ($R > 0.1$), which yielded 429 pairs. We then applied the Louvain community detection algorithm[106], which has demonstrated robust and efficient performance in comparative studies[172], to identify clusters of highly correlated cell states (Supplementary Fig 1a). This analysis produced five Louvain clusters (resolution = 1). The consensus and Louvain clusters showed strong overlap, and their shared cell states were defined as the core CEs (Fig. 4.1e and Supplementary Fig 1b-d). Remaining cell states that uniquely correlated with the core CE frequencies were incorporated to generate the extended CEs, encompassing most TME cell states (88.68%, 47 out of 53) (Fig. 4.1f and Supplementary Fig 1e).

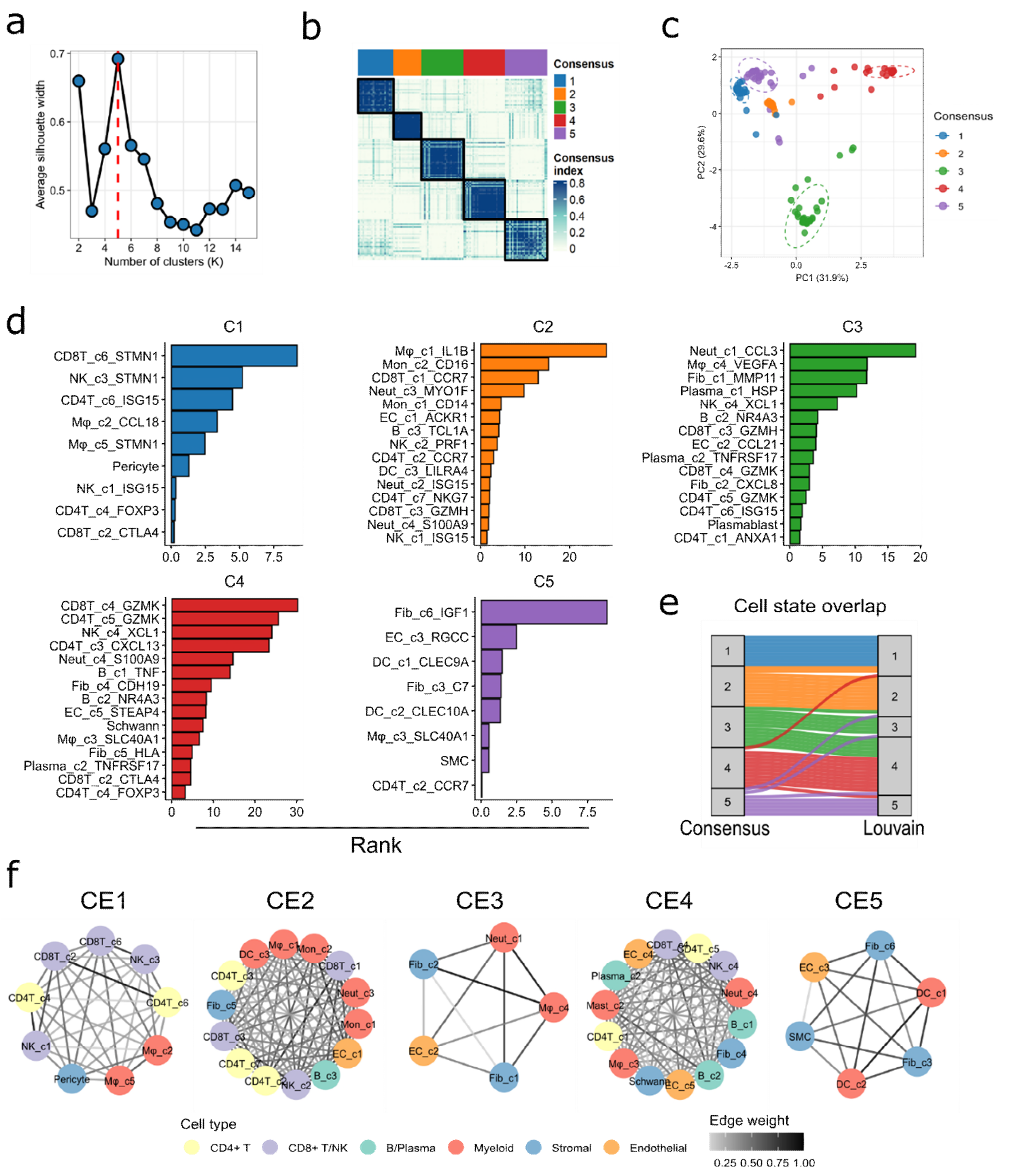


Figure 4.1 Cell state ranking and CE identification. (a) Optimal consensus cluster selection based on the ASW, with K = 5 identified as the optimal clustering number. **(b)** Heatmap of the consensus matrix, showing the five distinct consensus clusters. **(c)** PCA demonstrating distinct separation of consensus clusters in two-dimensional space. **(d)** Bar plots showing the top cell states ranking based on logFC x -

log10(p-value) for each consensus cluster. Statistical significance was calculated using a two-sided Wilcoxon rank-sum test. **(e)** Comparison of consensus clusters with Louvain clusters to identify overlapping cell states. **(f)** Network plots of the 5 CEs, with nodes representing cell states labelled with short names and coloured by major cell type. Edge colour indicates the Pearson correlation coefficient. PC: Principal component

4.3 Characterization of the identified CEs

Given that each patient sample exhibited a dominant CE, we assigned all samples into CETs based on their most abundant CE (Fig. 4.2a). Importantly, all CETs composed of multiple studies (> 3 studies), validating the robustness of the identified CEs (Fig. 4.2a-b).

CET frequencies significantly associated with tissue types (χ^2 test P = 9.5 x 10^{-6}) and neo-CTx (χ^2 test P = 6.5 x 10^{-5}) (Fig. 4.2a). Specifically, CE1, CE3, and CE4 abundance were significantly enriched in both PTs and MTs compared to NAT, while CE2 exhibited a moderate increase, though not reaching significance (Fig. 4.2c). Notably, CE3 was significantly enriched in MT compared to PT tissues (P = 0.013) (Fig. 4.2c). In contrast, CE5 was significantly enriched in NAT compared to PT and MT, while PTs also showed significant higher abundance compared to MTs (Fig. 4.2c).

We next compared CE frequencies between naïve and neo-CTx samples in primary tumour tissues. CE1, which composed of the immunosuppressive Mϕ_c2_CCL18 and dysfunctional CD8+ T (CD8T_c2_CTLA4) and NK cells (NK_c1_ISG15), was significantly depleted in neo-CTx samples (P = 0.007), followed by CE3 which almost reached significance (P = 0.054) (Fig. 4.2d). In contrast, CE4 which composed of activated lymphoid cells (CD8T_c4_GZMK, CD4T_c5_GZMK, NK_c4_XCL1) and neurotropic stromal cells (Fib_c4_CDH19 and Schwann cells) was significantly elevated in neo-CTx samples (P = 2.0 x 10^{-5}) (Fig. 4.2d). It is noteworthy that the CET4 group was also enriched with CE5 cell states, including complement fibroblasts (Fib_c3_C7), and antigen-presenting DCs (DC_c1_CLEC9A), suggesting the co-existence of these CEs (Supplementary Fig. 2a-b).

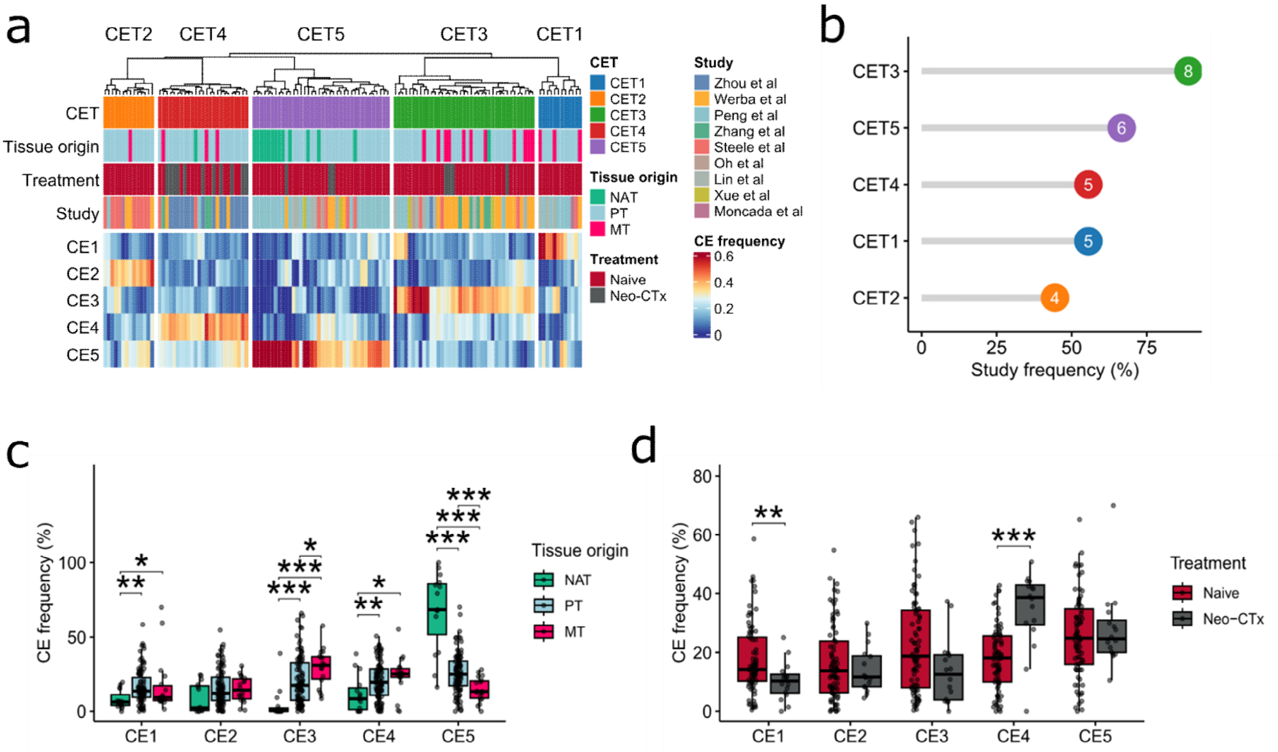
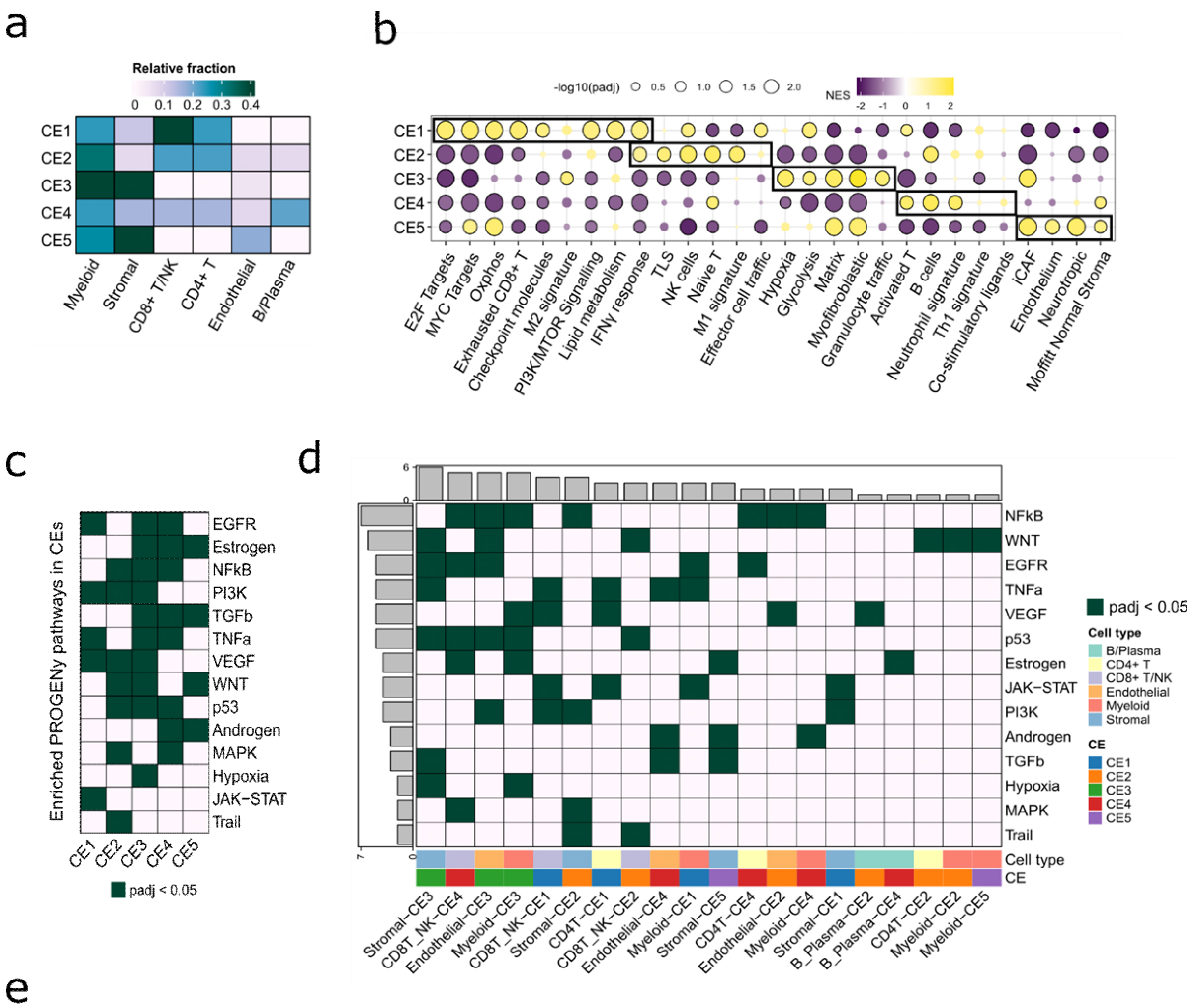


Figure 4.2 CE characteristics. (a) Heatmap showing the CE abundance across patient samples. Each sample is assigned a CET based on its dominant CE. **(b)** Study frequency in each CET. Number of studies is displayed in the dots. **(c)** Comparison of CE frequency across tissue types. Statistical significance was calculated using a two-sided Wilcoxon rank-sum test. *P<0.05, **P<0.01, ***P<0.001. **(d)** Comparison of CE frequency between naïve and Neo-CTx samples. Statistical significance was calculated using a two-sided Wilcoxon rank-sum test. *P<0.05, **P<0.01, ***P<0.001.

We then investigated the phenotypic profiles of CEs by performing gene signature analysis using curated and Hallmark gene signatures (Fig. 4.3a). CE1 was enriched with immunosuppressive (‘Exhausted CD8+ T’ and ‘Checkpoint molecules’) and proliferating gene sets (‘E2F Targets’, ‘MYC Targets’, and ‘PI3K/mTOR Signalling’). Notably, CE1 was the most metabolically active with enriched ‘Oxphos’, ‘Lipid metabolism’, and ‘Glycolysis’ (Fig. 4.2e). CE2 exhibited various anti-tumour immune gene sets, including ‘TLS’, ‘NK cells’, and ‘M1 signature’. Additionally, this CE was enriched with ‘naïve T’, as it consisted of naïve lymphocytes (CD8T_c1_CCR7, CD4T_c2_CCR7, and B_c3_TCL1A). CE3 which composed of fibroblasts (Fib_c1_MMP11 and Fib_c2_CXCL8) and hypoxic myeloid cells (Mϕ_c5_VEGFA and Neut_c1_CCL3), was enriched for ‘Hypoxia’, ‘Myofibroblastic’, and ‘Matrix remodelling’ gene sets, thus resembling a fibrotic/hypoxic associated phenotype. Furthermore, the neo-CTx enriched CE4, displayed elevated ‘Activated T’, ‘B cells’, and ‘Neutrophil signature’. Lastly, CE5 which composed of complement fibroblasts (Fib_c3_C7 and Fib_c6_IGF1) and endothelial cells (EC_c3_RGCC) was enriched for ‘Endothelium’, and ‘Neurotropic’ gene sets. Notably, both CE4 and CE5 were enriched for the ‘Moffitt Normal Stroma’, aligning with the enrichment of complement fibroblasts in both CEs (Supplementary Fig. 2a-b).

Next, we further validated the pathway activities of CEs and their components using PROGENy[173], which utilizes a large compendium of perturbation experiments to accurately infer pathway activities. Results show that most PROGENy pathways (64.3%, 9 out of 14) were shared in at least 3 CEs (Fig. 4.3c). However, certain pathways were unique to individual CEs, including JAK-STAT in CE1, Trail in CE2, and Hypoxia in CE3 (Fig. 4.3c). Specifically, both CE3 myeloid and stromal cells exhibited enriched Hypoxia activity while all CE1 components displayed enriched JAK-STAT activity. In CE2, both stromal and CD8+ T/NK cells showed enriched Trail activity. Taken together, these findings show diverse CEs with distinct tissue and treatment preferences, as well as phenotypic profiles (Fig. 4.3a and Fig. 4.3e).



e

CE	Description	Key cell types	Key cell states	PROGENy pathway	Tissue type	Neo-CTx
CE1	Immune-suppressive/Proliferative	CD8+ T/NK, CD4+ T, Myeloid	Mφ_c2_CCL18, CD4T_c4_FOXP3, CD8T_c2_CTLA4,CD8T_c6_STMN1,NK_c3_STMN1	JAK-STAT	PT/MT	Depleted
CE2	Naïve T/TLS	CD8+ T/NK, CD4+ T, B, Endothelial	Mφ_c1_IL1B, EC_c1_ACKR1, CD4T_c3_CXCL13, B_c3_TCL1A, CD8T_c1_CCR7, Fib_c5_HLA	Trail		
CE3	Hypoxic/Matrix remodelling	Stromal, Myeloid	Mφ_c4_VEGFA, Fib_c1_MMP11, Fib_c2_CXCL8, Neut_c1_CCL3	Hypoxia	PT/MT	Depleted
CE4	Activated T/Neo-CTx enriched	CD8+ T/NK, CD4+ T, B, Myeloid, Stromal	NK_c4_XCL1, CD8T_c4_GZMK, CD4T_c5_GZMK, Schwann, Mast_c2_LMNA, Fib_c4_CDH19		PT/MT	Enriched
CE5	Normal stroma/Vascularized	Stromal, Endothelial	Fib_c6_IGF1, Fib_c3_C7, EC_c3_RGCC, DC_c2_CLEC10A,DC_c1_CLEC9A		NAT	

Figure 4.3 CE pathway activities and annotations. (a) Heatmap showing cell type abundances in each CE. **(b)** Gene set enrichment analysis between CEs. Colour of dots represent the normalized enrichment score (NES) and the size represents the -log10 (p-value). Outlined dots represent significant differences (padj < 0.25). **(c)** Heatmap showing the enriched (padj < 0.05) PROGENy pathways in each CE. **(d)** Heatmap showing the enriched (padj < 0.05) PROGENy pathways across CE components grouped by major cell types. Bar plots on top and left represent the sum of columns and rows, respectively. **(e)** Description of each CE based on results from Fig. 4.2 and 4.3.

4.4 Spatial characterization of CEs

We then characterized CEs in spatial context, by mapping them onto spatial transcriptomics data (Visium) using CARD[108], as this deconvolution method is considered one of the top performers in a recent benchmarking study[174]. Our analysis showed clear coexistence of CEs in individual tumours with varying abundances. For instance, CE5 was the most abundant while the lymphocyte-enriched CE2 showed limited abundance (Fig. 4.4a). Results show significant spatial colocalization of cell states within CEs compared to cell states not belonging to the same CE (Fig. 4.4b-c). However, CE5 did not reach significance, possibly due to its high stromal coverage (Fig. 4.4c). A global colocalization score was then calculated for each CE across spatial sections (Methods 2.8.3). CE1 to CE3 displayed comparable colocalization levels, while CE4 and CE5 showed lower colocalization (Fig. 4.4d).

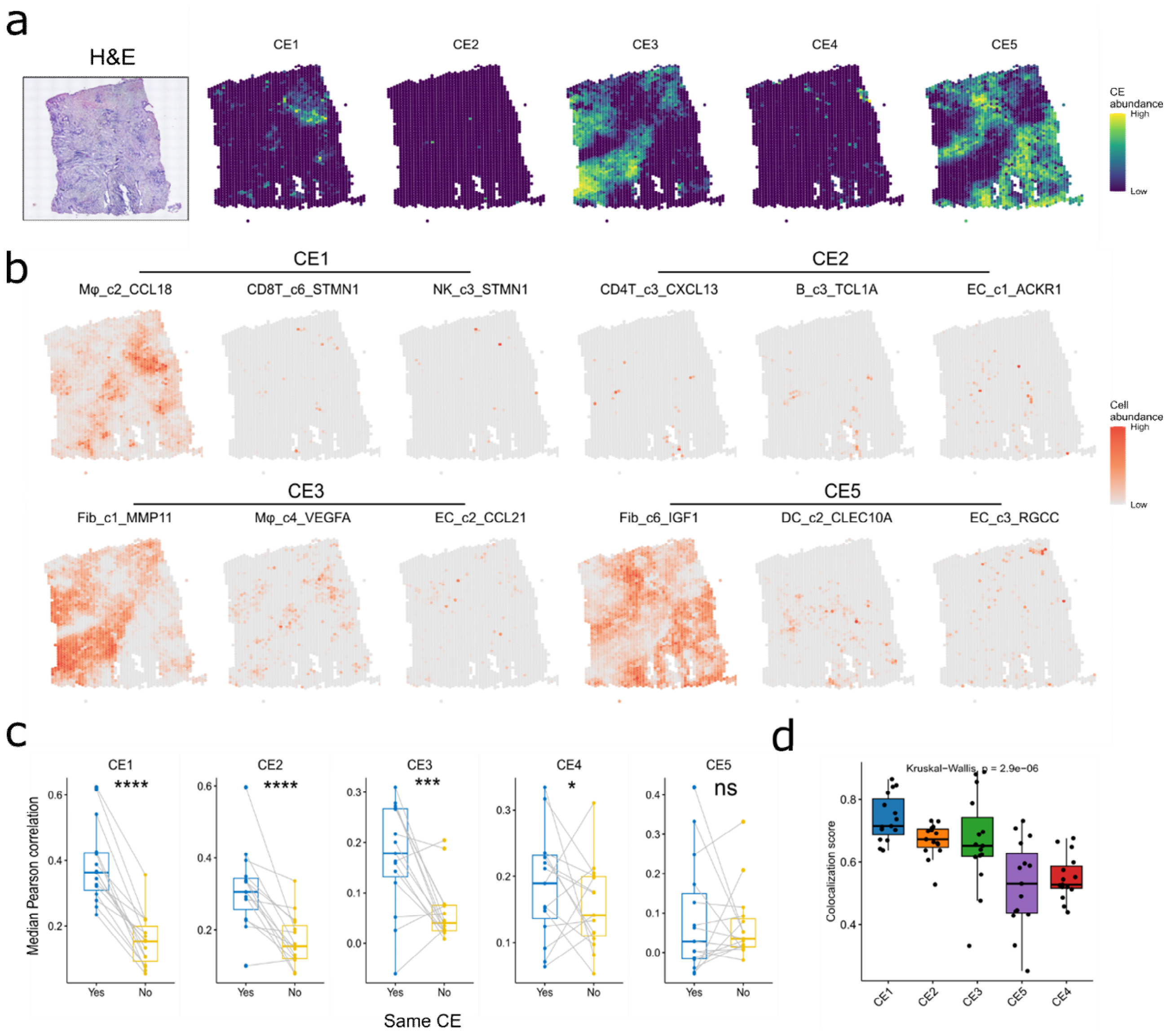


Figure 4.4 Spatial colocalization of CEs. (a) H&E staining, and spatial distribution of CE frequencies in a Visium PDAC section. **(b)** Spatial distribution of CE cell state frequencies in the Visium section shown in **a**. **(c)** Paired boxplots comparing the sample-based median Pearson correlation of cell states of different CEs and within the same CE. Statistical significance was calculated using a two-tailed paired Wilcoxon test. **(d)** Colocalization score of CEs. Statistical significance was calculated using Kruskal-Wallis test.

We then calculated the local Moran's I[109] as a measure of spatial autocorrelation, which quantifies the degree in which cell states within CEs aggregate in relation to their immediate neighbouring spots. Results show a prominent local aggregation of cell states within CEs compared to cell states of different CEs (Fig. 4.5a). Similar to the colocalization score, a global aggregation score was calculated for each CE (Methods 2.8.4). CE2, which resembles a compact TLS formation, displayed the highest aggregation score, while CE5 exhibited the lowest score as this CE was more spatially dispersed (Fig. 4.5b-d).

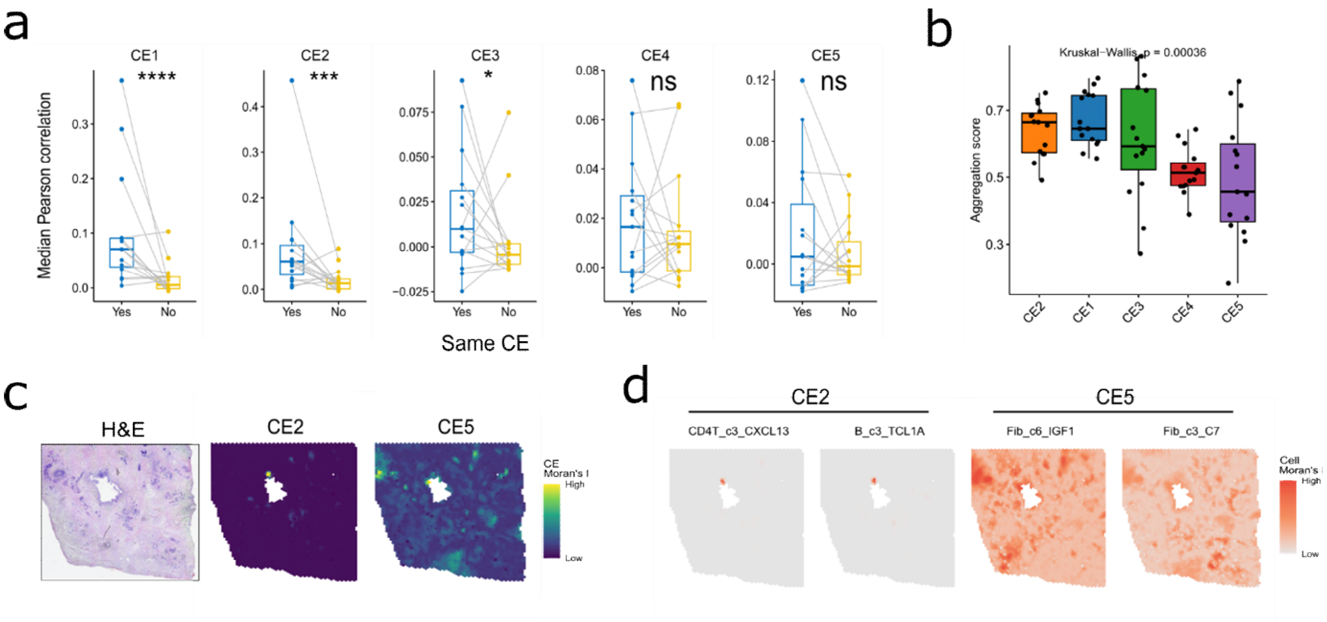


Figure 4.5 Spatial aggregation of CEs. (a) Paired boxplots comparing the sample-based median Pearson correlation of cell state aggregation of different CEs and within the same CE. Statistical significance was calculated using a two-tailed paired Wilcoxon test. **(b)** Aggregation scores of CEs. Statistical significance was calculated using Kruskal-Wallis test. **(c)** H&E staining, and spatial distribution of CE aggregations (local Moran's I) in a Visium PDAC section. **(d)** Spatial distribution of CE cell state aggregations (local Moran's I) in the Visium section shown in **c**.

Next, we sought to determine whether the CEs exhibited distinct morphological phenotypes. To do so, we validated CE organization by performing spatial clustering using stLearn[110], which integrates gene expression, tissue morphology, and physical distance for more accurate cellular categorization (Fig. 4.6a). In a representative spatial section, most CEs showed strong alignment with independent morphological (Morph) clusters (Fig. 4.6b). Notably, both CE1 and CE3 aligned with Morph C2, which was enriched for both pathways belonging to CE1 (‘M2 signature’ and ‘E2F Targets’) as well as CE5 (‘Myofibroblastic’ and ‘myCAF’) (Fig. 4.6b-c). This highlights the interconnection of CE1 and CE3 in the TME, which is supported by their co-occurrence in the scRNA-seq atlas (Supplementary Fig 1e). CE2 distinctly aligned with Morph C7, which displayed elevated CE2-specific gene signatures (‘B cells’, ‘TLS’, and ‘Effector cells’). Moreover, CE4 strongly aligned with Morph C1 which was enriched with CE5 gene signatures (‘Moffitt Normal Stroma’, ‘Endothelium’, and ‘Neurotropic’), consistent with the scRNA-seq results (Fig. 4.3b).

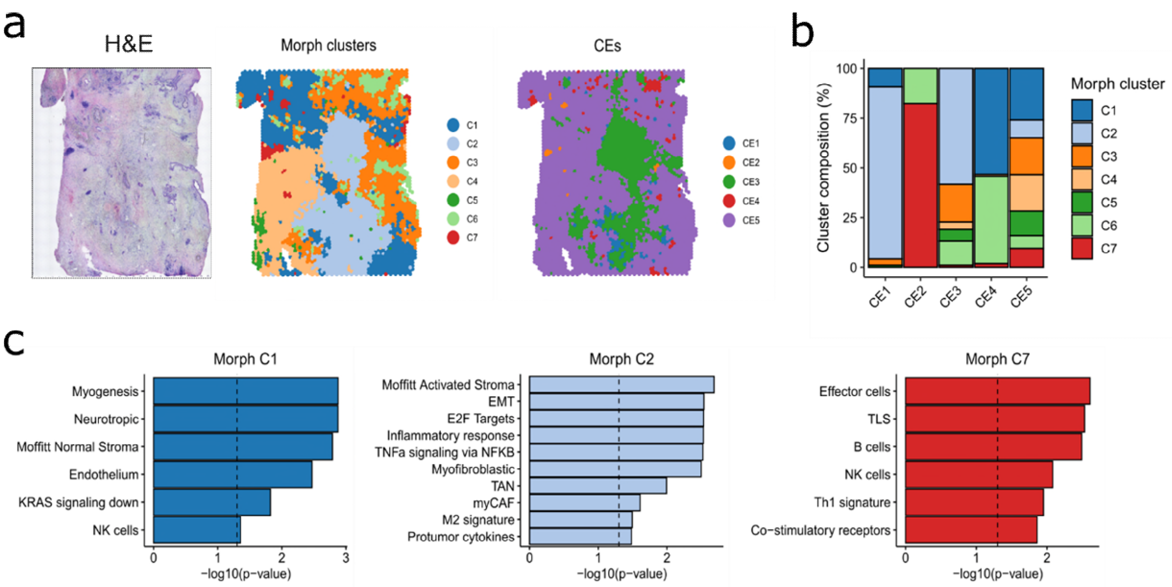


Figure 4.6 Comparison of CEs with Morph clusters. (a) H&E staining, and spot categorization based on Morph clustering (stLearn), and CE clustering. **(b)** Morph cluster composition in CE annotated spots. **(c)** Gene set enrichment analysis of curated and Hallmark gene signatures for Morph C1, C2, and C7. Gene sets displayed were filtered based on adjusted NES > 0 and p < 0.05. NES: Normalized Enrichment Score.

4.5 Cell-cell communications within CEs

CEs exhibit distinct communication networks essential for their formation and maintenance. To characterize these interactions, we analysed ligand-receptor-mediated cell-cell communications using CellChat[112]. Our results revealed that CE5 and CE3 produced the highest number of ligands across multiple categories (Fig. 4.7a-b). Endothelial cells and stromal cells were the predominant senders and receivers of communication signals, which are key components in CE5 and CE3 (Fig. 4.7c and Fig. 4.3a). In contrast, lymphocytes (T and B cells) exhibited the lowest communication signals. This may be due to their compact spatial organization, which facilitates more specific intercellular communications (Fig. 4.5a-c). Unlike lymphocyte-enriched CEs, stromal cells are more spatially dispersed and therefore produce a wide range of signalling molecules, enabling diverse interactions across longer distances[175]. Notably, the EC_c5_STEAP4 (CE4) displayed the highest number of outgoing signals compared to other CE members, followed by the two complement fibroblasts (Fib_c3_C7 and Fib_c6_IGF1) (Fig. 4.7d). Moreover, the immunosuppressive Mϕ_c2_CCL18 displayed the most outgoing signals compared to other immune cells, reflecting its critical role in shaping the TME.

We next investigated the relative communication pathway activities across CEs. We observed substantial differences among CEs composed of similar cell types. For example, the lymphocyte-enriched CE1 showed elevated activity in CD86, ApoE, and SIRP pathways, whereas CE2 was characterized by higher activity in CD34, CD23, and SELL pathways (Fig. 4.7e). This was also observed in stromal cell-enriched CEs. CE3 showed elevated SERPINE1, PLAU, and SPP1 angiogenic pathways, while CE5 showed higher activity in PDGFRB, ADGRG, and COMPLEMENT pathways. Notably, CE4 exhibited distinct activation of the KIT pathway, which is a key signalling axis for mast cell activation through *KITLG-KIT*[176].

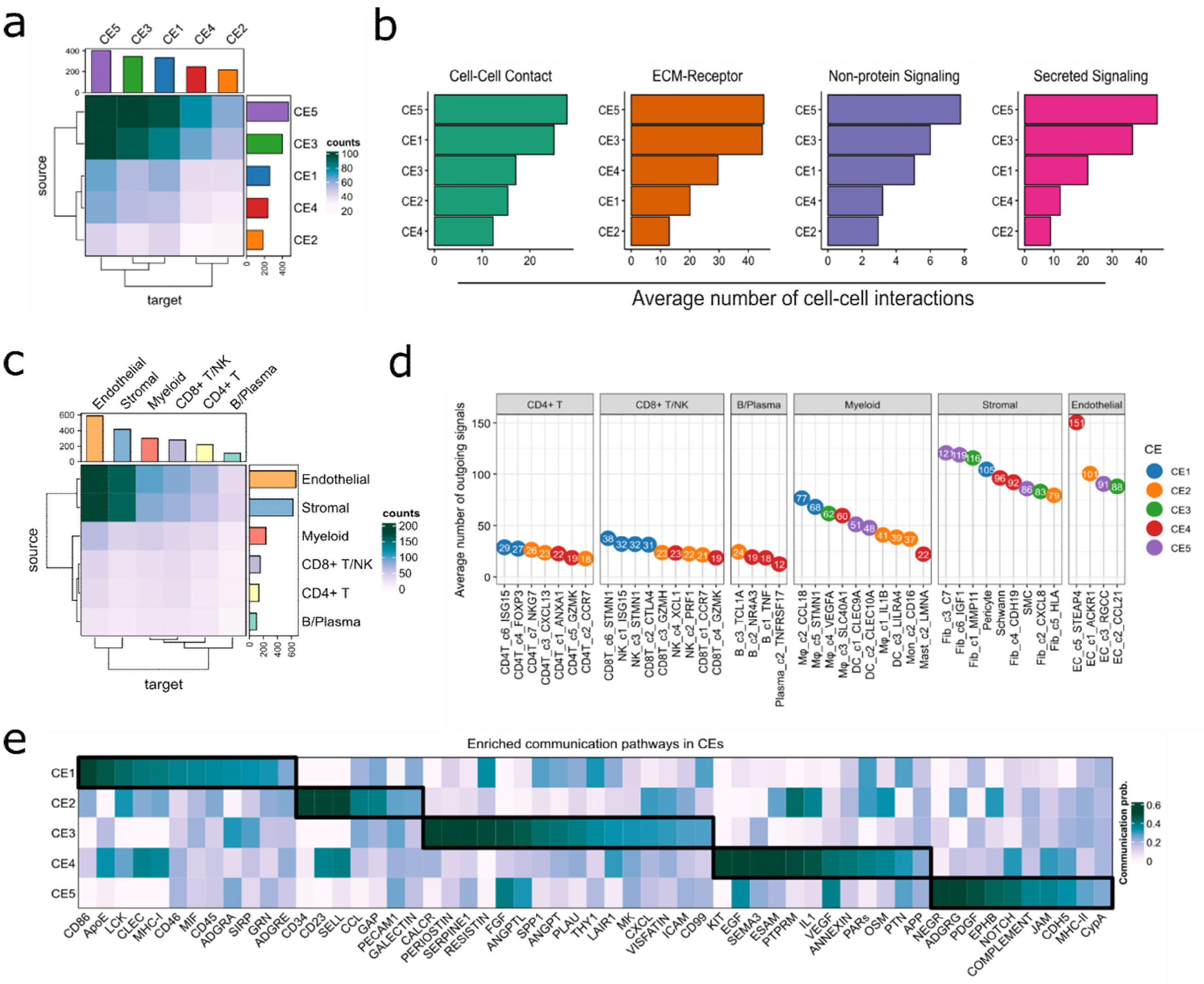


Figure 4.7 Cell-cell communications among CEs. (a) Heatmap showing the cell-cell interaction counts across CEs using CellChat. Bar plots on the top and right represent the sums of the columns and rows, respectively. **(b)** Average cell-cell interaction counts based on ligand-receptor pair types using CellChat. **(c)** Heatmap showing the secreted ligand-receptor pairs enriched (padj < 0.05) among CEs. Left bar shows the sum of rows and colours represent each contributing CE. **(d)** Dot plot showing the average number of outgoing signals across CE cell states facetted by major cell types. **(e)** Heatmap showing the relative communication probability across CEs.

We then investigated the specific communications between CEs, by utilizing the CellChat database of L-R pairs. The majority of differentially enriched L-R pairs were unique to an individual CE (65.4%, 286 out of 437). The CXCL14-CXCR4 was the most common signalling axis, and was shared across four CEs (CE2-5), while SPP1-CD44 was shared in three CEs (CE1, CE3, CE4) (Fig. 4.8a). Additionally, CCL3-CCR1 signalling was enriched in both CE1 and CE3, suggesting an important role in macrophage recruitment and M2-polarization, which are key features of both CEs[177]. Notably, the majority of checkpoint molecules, including CD86-CTLA4, NECTIN2-TIGIT, LAIR1-LILRB4, and LGALGS3-LAG3, were specifically enriched in CE1 and originated mainly from Myeloid-CE1 (Mϕ_c2_CCL18) towards CE1 CD8+ T/NK cells (Fig. 4.8a-b). Additionally, this CE was specifically enriched for diverse chemokine axes mediating macrophage recruitment (CCL5/7/13/23-CCR1 and CSF1-CSF1R), as well as Treg recruitment and activation (CCL18-CCR8 and CCL20-CCR6).

CE2 exhibited elevated proinflammatory TNF-TNFRSF1B directed towards CE2 myeloid cells, as well as CE1 T and NK cells, indicating possible inter-CE communications. Additionally, CE2 was enriched for IL16-CD4 signalling, which is reported to drive Th1 cell polarization and associated with favourable responses to ICB immunotherapy[178]. Notably, CE2 uniquely displayed elevated BTLA-TNFRSF14 checkpoint signalling originating from CE2 CD4+ T cells towards B and CE2 CD8+T/NK cells[179], highlighting distinct immunosuppressive mechanisms across CEs.

Further, multiple stromal cells were enriched for the pleiotropic growth factor *MDK*, which interacted with several receptors (Fig. 4.7e and 4.8b)[180]. However, CE3 uniquely exhibited MDK-SDC1 signalling, as *SDC1* enrichment was restricted to CE3 stromal cells. Therefore, while certain ligands are shared across CEs, individual CEs could engage in distinct signalling pathways. In addition, CE3 displayed elevated SERPINE1-PLAUR signalling originating from CE3 stromal cells targeting both CE3 stromal and CE3 myeloid cells. Beyond its pro-fibrotic functions[181], this cytokine also promotes M2 macrophage polarization, thus driving immunosuppression[182].

CE4 and CE5 shared multiple growth factor signalling including AREG-EGFR, HBEGF-EGFR, and PDGFRD-PDGFRB, as well as the chemokine CXCL12-CXCR4 signalling axis. Consistent with the pathway analysis using CellChat, CE4 was also enriched for the KITLG-KIT from CE4 endothelial cells towards CE4 myeloid cells, indicating mast cell activation (Fig. 4.7e and Fig. 4.8b). CE5 was uniquely enriched for PDGFRB-PDGFRA/B signalling, originating from CE5 endothelial cells and targeting CE5 stromal cells. This pathway is active during normal tissue conditions[183], which is a hallmark for this CE.

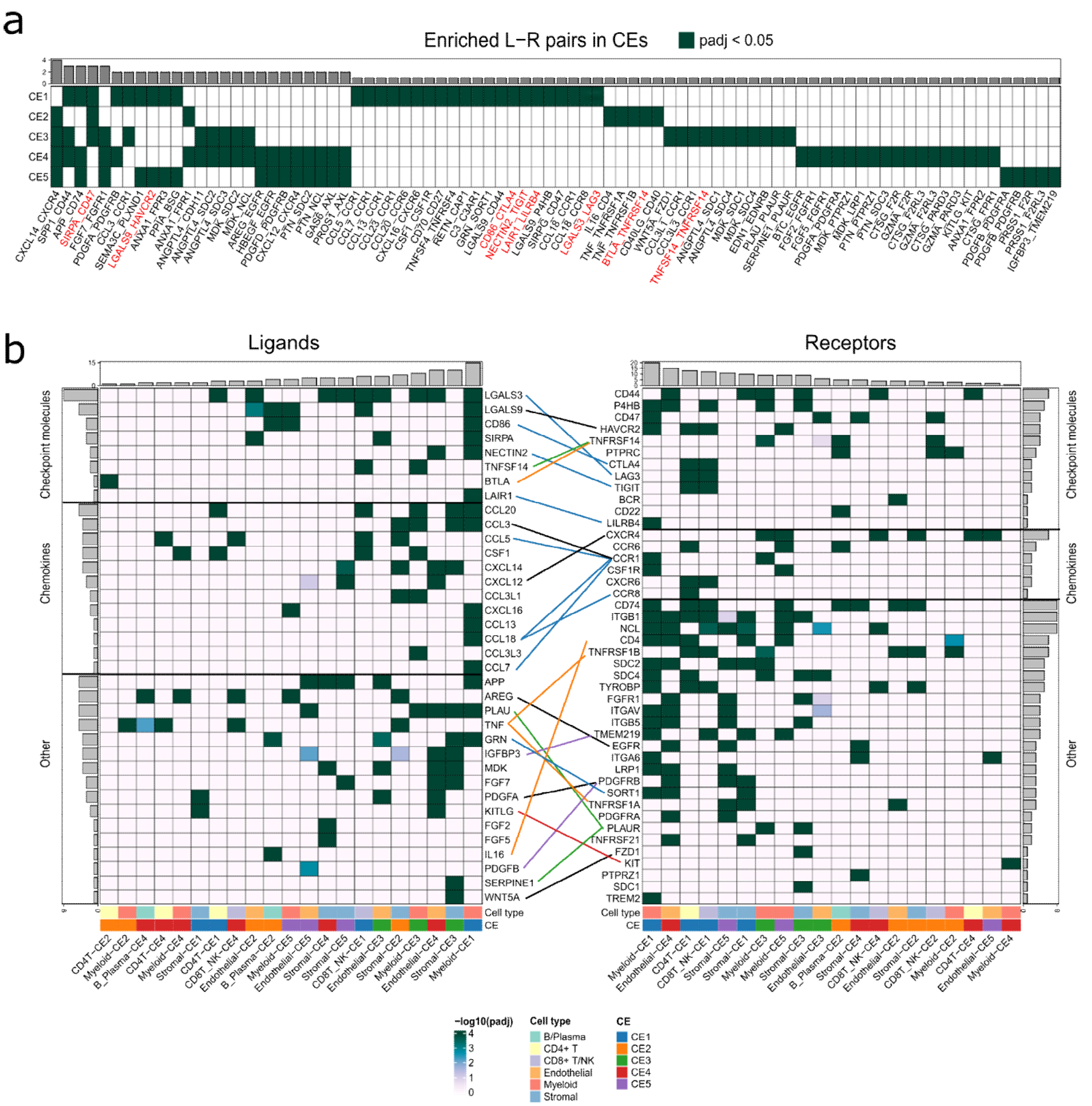


Figure 4.8 CE-specific ligand receptor interactions. (a) Heatmap showing the enriched ligand-receptor pairs across CEs. Top bar plot represents the sum of the columns. Checkpoint molecules are coloured by red. Statistical significance was calculated using an unpaired Wilcoxon test. **(b)** Heatmaps of enriched ligands (left) and corresponding receptors (right) across CE cell type members. Bar plots on the top and left/right represent the sum of the columns and rows, respectively. Bottom annotation is coloured based on the cell type and CE group. Links connecting the ligand from left heatmap and corresponding receptor on right heatmap are coloured based on their unique CE. Links are coloured black if they are active in more than one CEs.

4.6 Conclusions and Discussion

Diverse cell types communicate with each other to form cellular niches that drive cancer progression and therapeutic resistance. However, understanding the key cell types and the mechanisms by which they form these cellular niches is largely unresolved. In our analysis we identified five robust CEs with distinct phenotypic, spatial organization, and cell-cell communication profiles. Unlike the Puleo et al five TME subtype classification which relied on broad gene signatures in bulk RNA-seq datasets[116], we leveraged the single-cell atlas to identify more accurate subtypes based on ground truth proportions. More recently, Sivakumar et al characterized myeloid-enriched (ME) and adaptive-enriched (AE) groups in PDAC using multi-omics approaches, uncovering key differences in signalling mechanisms[184]. In contrast, we identified sub-categories within the ME group, with CE1 and CE3 both enriched in macrophages but exhibiting marked differences in functional, and metabolic activities as well as intercellular communications.

Moreover, our CE analysis further validates the critical role of Mϕ_c2_CCL18 in driving PDAC immunosuppression, potentially through the establishment of CE1 via Treg recruitment and regulation of CD8+ T and NK cell activity. This suggests that only a subset of patients with elevated CE1 abundance (CET1) may selectively benefit from ICB therapy by reversing T cell dysfunction. Differential ligand-receptor analysis revealed potential CE1-specific ICB therapeutic targets, including CD86-CTLA4, NECTIN2-TIGIT, and LGALS9-HAVCR2, whereas PD-L1-PD1 was not significantly enriched. Notably, phase II clinical trials of CTLA-4 monotherapy in locally advanced or metastatic PDAC reported poor response rates[185 186]. It is possible that most patients enrolled in these studies belonged to other CETs that are potentially refractory to ICB responses. Importantly, nearly all patients in these studies received standard-of-care chemotherapy prior to ICB. Given that neo-CTx reduces CE1 abundance, this may help explain the lack of therapeutic efficacy.

PDAC is mainly composed of a dense fibrotic stroma that impairs vascularization and limits immune cell infiltration[187]. The hypoxic CE3, composed of Mϕ_c4_VEGFA, and Fib_c1_MMP1/Fib_c2_CXCL8 populations, aligns with recent studies reporting the co-occurrence of SPP1+ macrophages and FAP+ fibroblasts, which collectively restrict T cell infiltration and drive resistance to ICB therapy[188 189]. Notably, CE3 correlated with the immunosuppressive CE1 in both co-abundances, and spatial colocalization. This suggests that $CE1^{high}CE3^{high}$ PDAC tumours might be refractory to ICB therapy compared to $CE1^{high}CE3^{low}$ tumours, due to CE3-driven immune exclusion surrounding the tumour compartment. Therefore, combination strategies targeting vulnerabilities in both CEs may yield more effective responses. For instance, a recent phase II study (TM5164-MM trial) in advanced melanoma demonstrated that combining a PAI-1 inhibitor (TM5614) with anti-PD-1 antibody was both well tolerated and effective[190]. Collectively, these findings provide key insights into the PDAC TME heterogeneity with potential clinical relevance. Further analyses of bulk RNA-seq data will be performed in later chapters to validate the association of these CEs with clinical outcomes and treatment responses

Chapter 5

Associations of CEs with clinical outcomes

5.1 Introduction and Aims

Due to the late onset of symptoms and lack of adequate screening methods, the prognosis of PDAC remains poor, with less than 20% of patients presenting with resectable disease at the time of diagnosis[191]. Even after primary tumour resection, most patients develop early disease recurrence with 5-year survival rate of 12-17%[192]. Furthermore, due to the lack of effective surveillance for improving post-operative survival[192], more accurate biomarkers for prognosis and predicting recurrence are urgently needed. Of note, the carbohydrate 19-9 (CA19-9) is the most extensively studied biomarker for prognostic surveillance in PDAC, and a relative postoperative increase in CA19-9 has been shown to precede detection of recurrence for up to 7 to 10 months[193]. However, several limitations which impair the value of this biomarker have been identified. For example, approximately 6% of Caucasian population is Lewis-antigen negative, leading to reduced CA19-9 production[194]. This poses a significant clinical challenge as patients with low or undetectable CA19-9 levels may receive suboptimal chemotherapy, as well as delayed recurrence detection[194]. Additionally, CA19-9 have been reported to not sufficiently distinguish between malignant and benign obstructive jaundice diseases[195]. What is more, this antigen has displayed a low sensitivity and specificity of 70-74% for distinguishing PDAC from healthy and chronic pancreatitis groups[196], further emphasizing the need for more accurate prognostic biomarkers.

Currently, patients with locally advanced or metastatic PDAC receive standard of care FOLFIRINOX or gemcitabine/paclitaxel combination chemotherapy. Despite its improved therapeutic benefit compared to gemcitabine, responses remain heterogeneous and most patients are resistant to this treatment regime[14]. This variability can be attributed to the substantial inter-tumour heterogeneity evident in PDAC[197]. Findings from the COMPASS trial further demonstrated that basal-like tumours are refractory to FOLFIRINOX[198], highlighting the potential utility of molecular classification for treatment selection. However, recent scRNA-seq studies revealed extensive ITH, with basal-like and classical cancer cells frequently co-existing in individual tumours[107]. This poses a major challenge to treatment responses and complicates the clinical application of subtype-based patient stratification.

Moreover, advancements in scRNA-seq technology have provided deeper insights into the TME, uncovering cell types with prognostic relevance across multiple cancer types. For instance, the *CXCL9/SPP1* ratio, which reflects macrophage polarity distinct to the conventional M1/M2 classification, has shown strong prognostic value in head and neck squamous cell carcinoma (HNSC)[199]. In PDAC, TME-derived biomarkers such as the neutrophil/lymphocyte ratio and transcriptomic immunophenotypes have similarly been associated with patient survival[200]. However, these approaches do not capture the extensive cellular diversity within the TME or the combined impact of cellular ecosystems on patient prognosis and treatment responses.

Therefore, in this chapter, we aim to investigate the prognostic and predictive significance of the identified cell states and CEs by utilizing multiple bulk RNA-seq datasets with clinical information. The specific aims of this chapter include:

- Develop a deconvolution tool that would enable accurate estimation of cell type and cell state abundances in PDAC and benchmark its performance across other deconvolution methods.
- Investigate the association of TME cell states and CEs with patient survival and metastatic potential.
- Investigate the association of TME cell states and CEs with chemotherapy and immunotherapy responses.
- Validate CE biomarkers for predicting patient survival using patient-derived samples.

5.2 Robust cell type and cell state deconvolution of bulk transcriptomic data

Using our large-scale atlas, we developed a deconvolution tool for both cell types (CTcaller) and cell states (CScaller), designed to accurately estimate cell abundances in bulk RNA-seq datasets. Briefly, we obtained refined signature genes whose expression correlated with ground-truth abundances of the corresponding cell types or states (Methods 2.10.2). Next, for each bulk sample, average gene signatures of cell types were computed and normalized to yield probability distributions ranging from 0 to 1 within each sample.

We then compared this deconvolution framework with other reference-based deconvolution tools, including CIBERSORTx[121], InstaPrism[122] (a fast implementation of BayesPrism[201]), MuSiC[123], NNLS[123], and BisqueRNA[124]. We evaluated the estimated cell-type proportions from each deconvolution method against ground-truth proportions using both Pearson's and Spearman's correlations, and RMSE metrics. In the pooled analysis, CIBERSORTx achieved the best performance (R = 0.8, RMSE = 0.07), followed closely by our CTcaller (R = 0.74, RMSE = 0.09)

(Fig. 5.1a). To obtain more robust assessments, we next calculated the aforementioned metrics for each cell type individually. Both CTcaller and CIBERSORT showed the highest median Pearson's and Spearman's correlations, and 1-RMSE scores, compared with other deconvolution tools (Fig. 5.1b). We then calculated the percentage of cell types with significant correlation between the estimated and ground truth proportions across methods. CTcaller outperformed other methods with all cell types showing significant correlations, followed by CIBERSORTx, and InstaPrism (Fig. 5.1c).

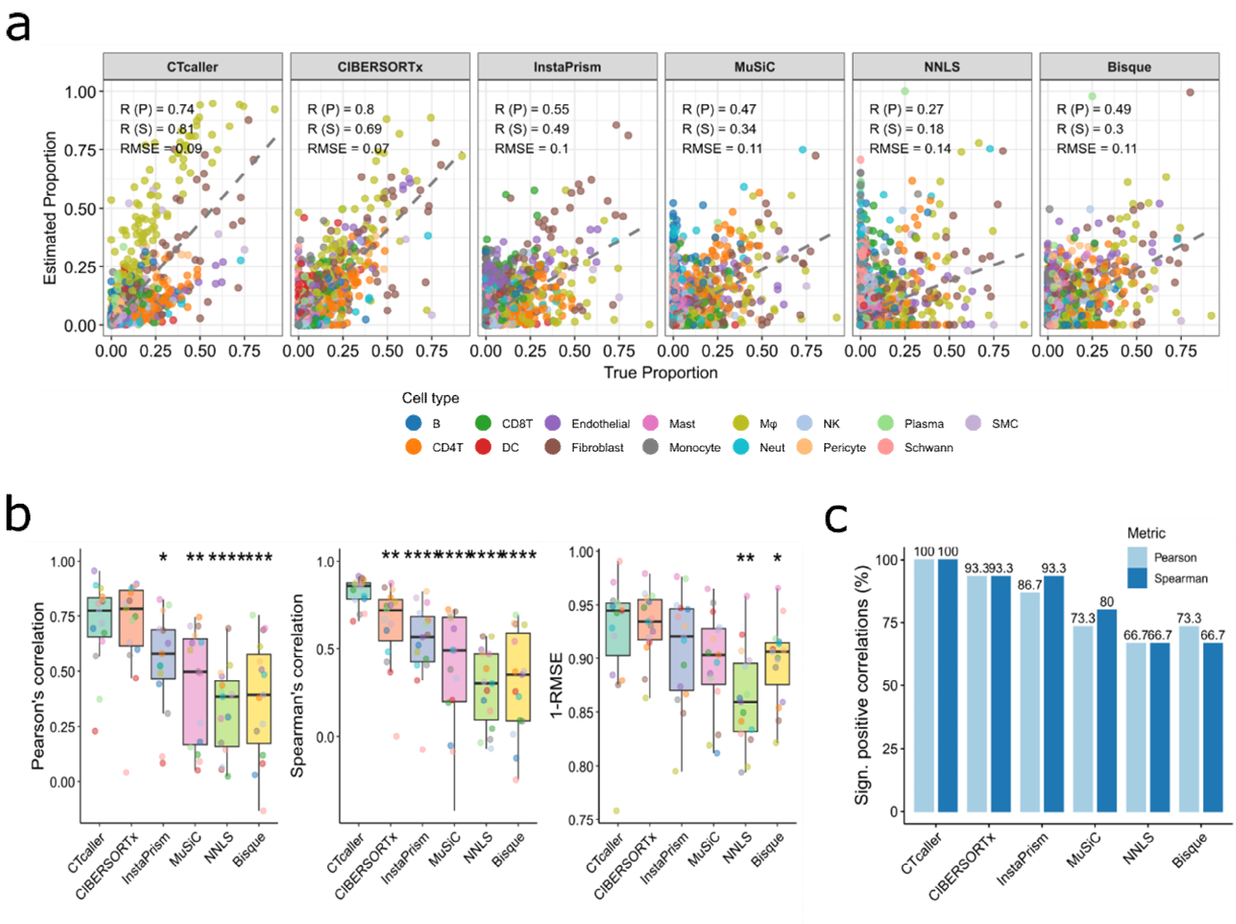


Figure 5.1 Cell type deconvolution benchmarking. (a) Correlation of estimated cell type abundances with ground truth proportions facetted by deconvolution methods. Points are coloured based on the cell type. Both the Pearson's (P) and Spearman's (S) correlation coefficients are displayed along with the RMSE. **(b)** Boxplots comparing the cell type-based performance metrics across deconvolution methods.

Statistical significance was calculated using a two-sided Wilcoxon rank-sum test using the CTcaller as reference. *P<0.05, **P<0.01, ***P<0.001. **(c)** Bar plots showing the frequency of cell types with significant positive correlations ($R > 0$ and $P < 0.05$) between their ground truth proportions and estimated proportions. RMSE: Root mean squared error.

The same analysis was then repeated for all 53 TME cell states. Similar to the cell type level, CIBERSORTx achieved the best performance (R = 0.5, RMSE = 0.05) in the pooled analysis, followed closely by CScaller (R = 0.44, RMSE = 0.05) (Fig. 5.2a). However, when evaluating the deconvolution accuracy for each cell state individually, the CScaller significantly outperformed other tools with the highest median for all metrics (Fig. 5.2b-c).

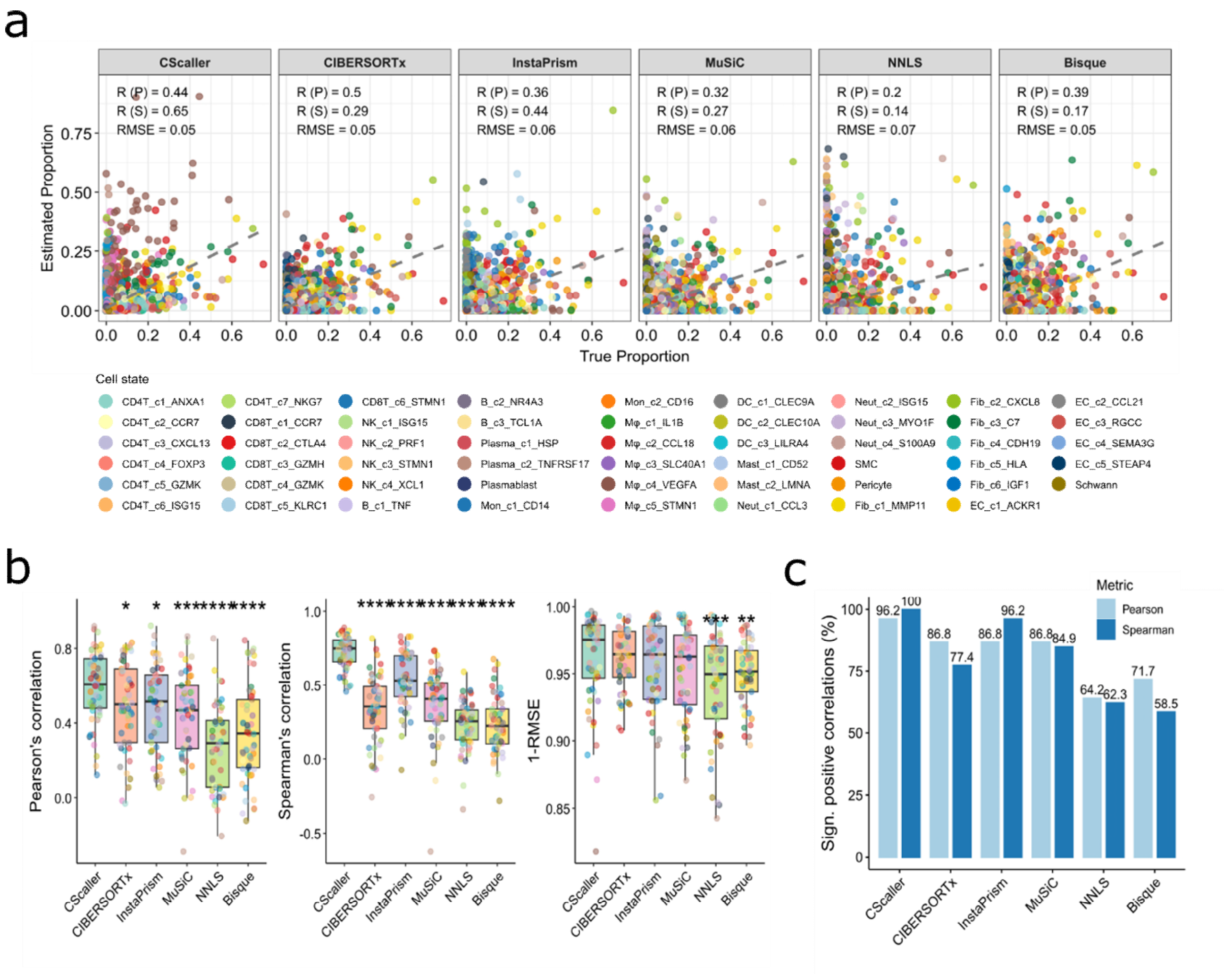


Figure 5.2 Cell type deconvolution benchmarking. (a) Correlation of estimated cell state abundances with ground truth proportions facetted by deconvolution methods. Points are coloured based on the cell state. Both the Pearson's (P) and Spearman's (S) correlation coefficients are displayed along with the RMSE. **(b)** Boxplots comparing the cell state-based performance metrics across deconvolution methods. Statistical significance was calculated using a two-sided Wilcoxon rank-sum test using the CScaller as reference. **P*<0.05, ***P*<0.01, ****P*<0.001. **(c)** Bar plots showing the frequency of cell states with significant positive correlations (R > 0 and *P* < 0.05) between their ground truth proportions and estimated proportions. RMSE: Root mean squared error.

To further evaluate the accuracy of our deconvolution framework, we performed the same benchmarking analysis in an independent single-nucleus RNA-seq (snRNA-seq) dataset from

Hwang et al[61], using their pre-annotated labels (Supplementary Fig. 3). CTcaller achieved consistently high correlations, comparable to other deconvolution methods, and achieved the highest median Spearman's correlation (Supplementary Fig. 3c-e). Collectively, these results highlight the utility of CTcaller and CScaller for accurately estimating TME cell type and cell state abundances, respectively, and for enabling downstream analyses in bulk RNA-seq datasets.

However, the strong performance of CTcaller and CScaller may be reflect optimization to the specific reference atlas, and may lack generalisability to independent cohorts. Although we validated performance using an external snRNA-seq dataset, future work incorporating blinded datasets from additional platforms and tumour contexts will be important for further assessing robustness.

5.3 Association of TME cell states and patient survival

We then utilized our deconvolution tool to decipher the prognostic landscape of TME cell states in resectable PDAC tumours from five bulk RNA-seq datasets (Supplementary Table 12). For each individual study, we estimated TME cell state abundances using CScaller, and performed multivariate Cox regression to account for age, sex, and stage (Fig. 5.3a). Prognostic genotypes, such as *KRAS*^{WT} or HRR deficiency, were not available as control variables in our analysis. Results show substantial heterogeneity in overall survival between cell states belonging to the same cell type. For instance, the Fib_c2_CXCL8 and Fib_c1_MMP11 correlated the strongest with poor survival while Fib_c3_C7 and Fib_c6_IGF1 correlated with favourable survival (Fig. 5.3a and 5.3c). In line with previous reports, the SPP1^{+} M2-like macrophage correlated strongly with poor outcomes across cohorts[188]. Interestingly, the CCL21^{+} endothelial cell population (EC_c2_CCL21) was associated with poor outcomes compared to other endothelial cell states, including EC_c1_ACKR1, and EC_c4_SEMA3G, suggesting opposing roles regarding disease progression. Notably, the association between cell state abundance and patient outcomes was consistent across the five cohorts for both overall survival and disease-free survival, despite differing sequencing platforms used between studies (Fig. 5.3b, Supplementary Fig. 4a and Supplementary Fig. 4c). To further validate these results, we performed the survival analysis using InstaPrism deconvolution (Supplementary Fig. 4b). Association between cell state abundance and patient outcomes was consistent between the two deconvolution methods, with positive correlations in four bulk RNA-seq cohorts, except for TCGA (Supplementary Fig. 4d).

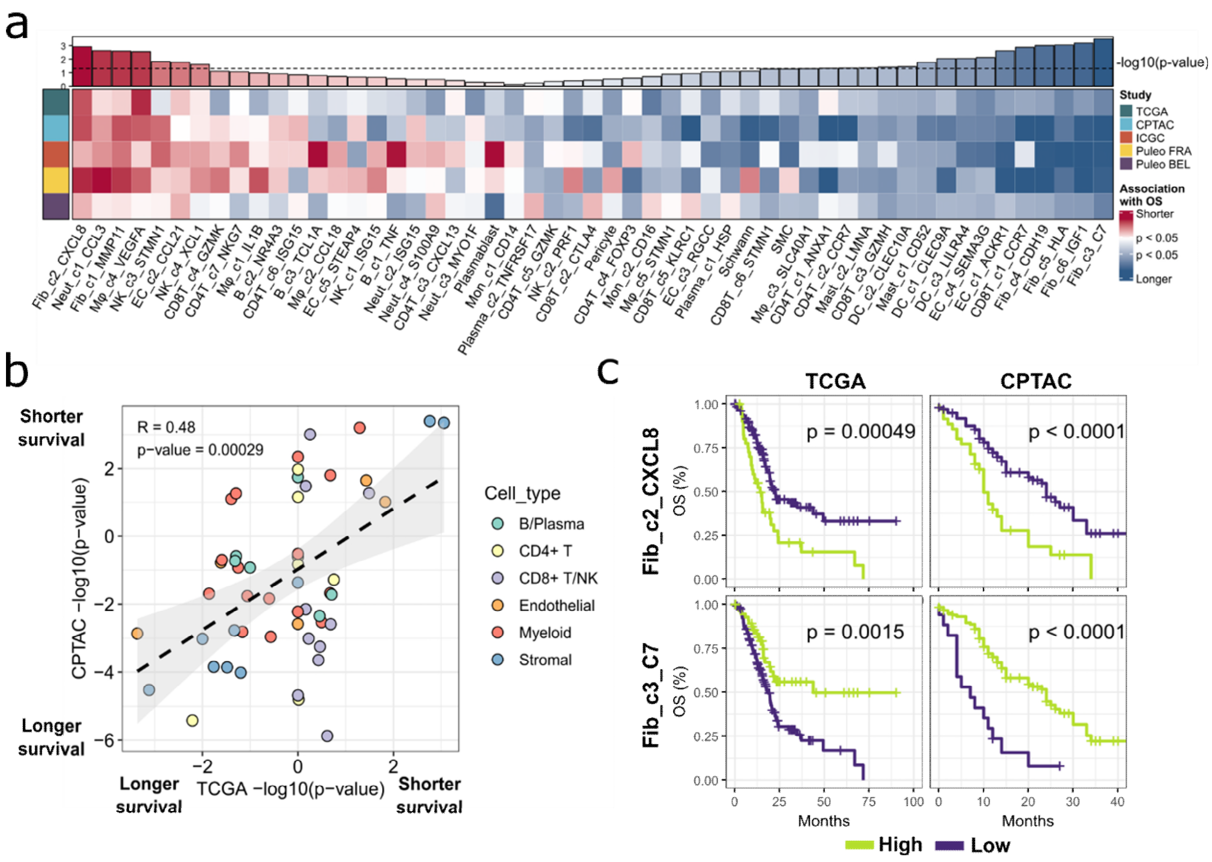


Figure 5.3 Prognostic landscape of TME cell states. (a) Heatmap showing the association of cell states with overall survival (OS) through multivariate Cox regression modelling. If cell states are associated with worse outcomes (HR > 1), the survival association is presented as -log10 (p-value). If cell states are

associated with improved outcomes (HR < 1), the survival association is presented as -log10 (p-value) multiplied by -1. Bar plot on top represents the average the survival association for each cell state. **(b)** Correlation of cell state survival associations between the CPTAC and TCGA cohorts. Points are coloured based on cell type. Statistical significance was calculated using Pearson's correlation. **(c)** Kaplan-Meier survival curves for the fibroblasts Fib_c2_CXCL8 and Fib_c3_C7 in both the TCGA and CPTAC cohorts.

5.4 Association of CEs with patient survival and metastasis

We then estimated CE abundances in the bulk datasets and stratified patients into CETs, based on their most abundant CE (Methods 2.10.4). Similar to the single-cell atlas, CET5 was distinctly enriched in NAT tissue samples (χ^2 test $P = 2.7 \times 10^{-11}$) (Fig. 5.4a). CET frequencies were comparable across datasets, including the single-cell atlas, with the CET3 being the most frequent (36.2%, 281 out of 776), while the neo-CTx-enriched CET4, was the least frequent (7.1%, 55 out of 776) (Fig. 5.4b). The TLS-like CET2, was found in 11.3% of patients (88 out of 776), similar to the reported frequencies of intratumoural TLS in PDAC (15.9%)[202]. Kaplan-Meier survival analysis showed significant associations of the identified CETs with both overall (OS) and disease-free survival (DFS) (Fig. 5.4c). CET5 displayed the longest median survival (OS: 42 months, DFS: 31 months), followed by CET2 (OS: 32 months, DFS: 24 months). In contrast, CET3 was linked with worst survival displaying the shortest median survival (OS: 16 months, DFS: 10 months), followed by CET1 (OS: 20 months, DFS: 16 months). Multivariate Cox regression showed that higher CE3 abundance was consistently associated with worse survival across studies, while CE5 was associated with favourable survival (Fig. 5.4d).

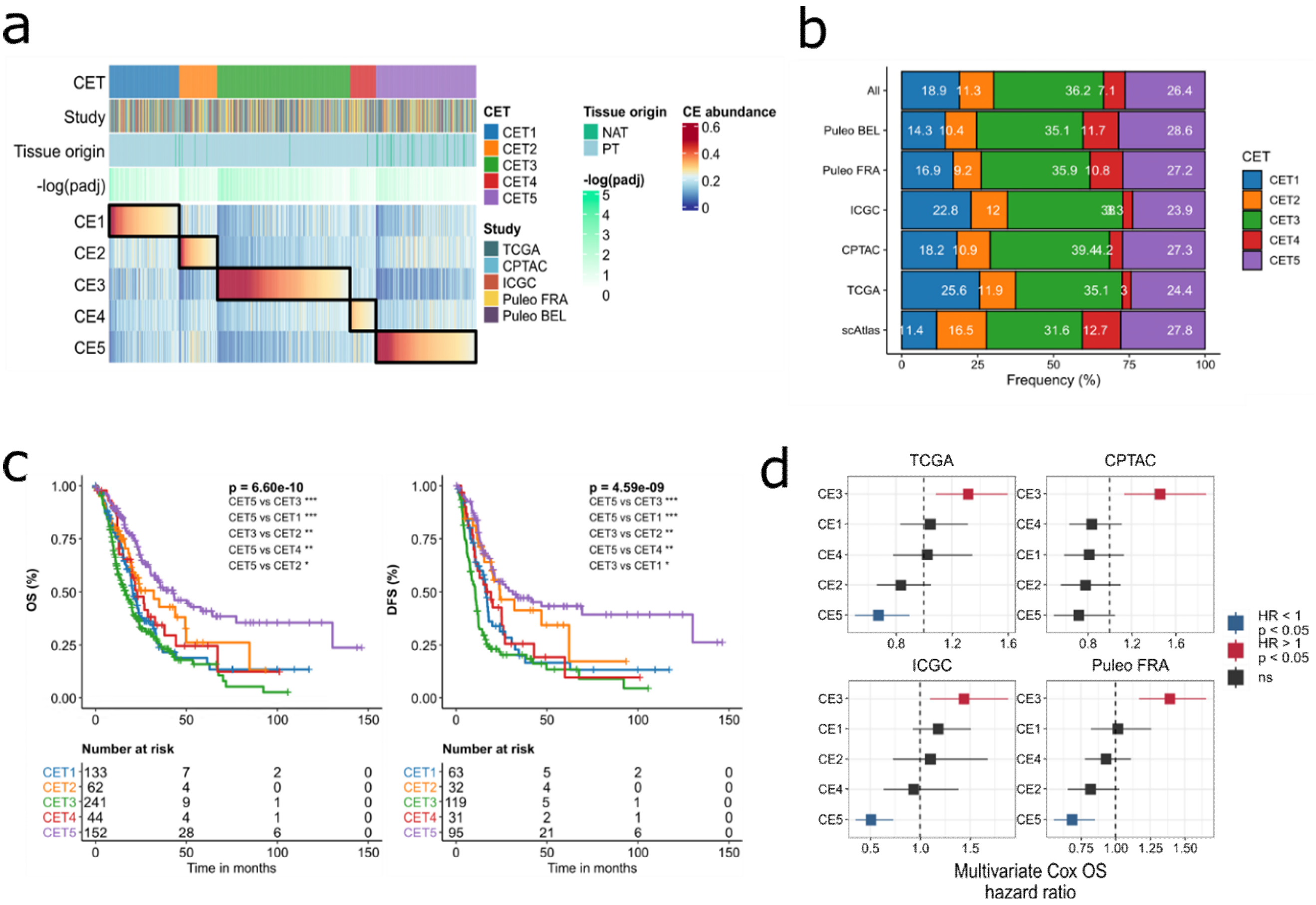


Figure 5.4 Association of CEs with patient outcomes and metastases. (a) Heatmap showing the CE abundances across the five PDAC cohorts, grouped into CETs. **(b)** CET frequencies within individual bulk cohorts, the scAtlas, and combined datasets. **(c)** Kaplan-Meier survival curves for the assigned CETs using all PDAC cohorts for both OS (left) and DFS (right). Statistical significance was determined using the two-sided log-rank test for each CET pair. **(d)** Forest plots showing the association of CE abundances as a continuous variable with overall survival using multivariate Cox proportional hazards regression.

We then correlated CE abundances with different types of metastatic spread. Only the CPTAC cohort included data for the presence of vascular or perineural invasion, and was therefore used for analysis. CEs did not significantly differ in vascular invasion, however, both CE4 (*P* = 0.002) and CE5 (*P* = 0.04) were significantly associated with the presence of perineural invasion (Fig. 5.5a). This metastatic potential aligns with the cellular composition of both CEs, as they are distinctly enriched with neurotropic fibroblasts and Schwann cells. Next, we utilized the recent study from Link et al[117], that compared metastatic PDAC that either spread to the lung (n = 25) or the liver (n = 85), with the former group associated with significantly longer survival. Our results show that both CE2 (*P* = 0.048) and CE5 (*P* = 0.027) were enriched in the lung cohort, while the more aggressive CE3, was enriched in the liver cohort (*P* = 0.02) (Fig. 5.5b). Notably, CE5 abundance was independently associated with longer survival in both the liver, and lung cohorts (Fig. 5.5b). Next, we utilized our in-house microarray cohort (Oxford-Kidani) that included information of distant (predominantly liver) or local (pancreas and regional lymph nodes) and distant metastases at the time of tumour resection. Results show that the immunosuppressive CE1, was significantly associated with distant metastases (*P* = 0.01), and to a lesser extent with both local and distant metastases (*P* = 0.049) (Fig. 5.5c). Conversely, CE5 was inversely associated with local and distant metastases (*P*= 0.049). Taken together, these findings reveal diverse associations of CEs with clinical outcomes and metastatic potential, suggesting potential utility of this classification system for patient stratification.

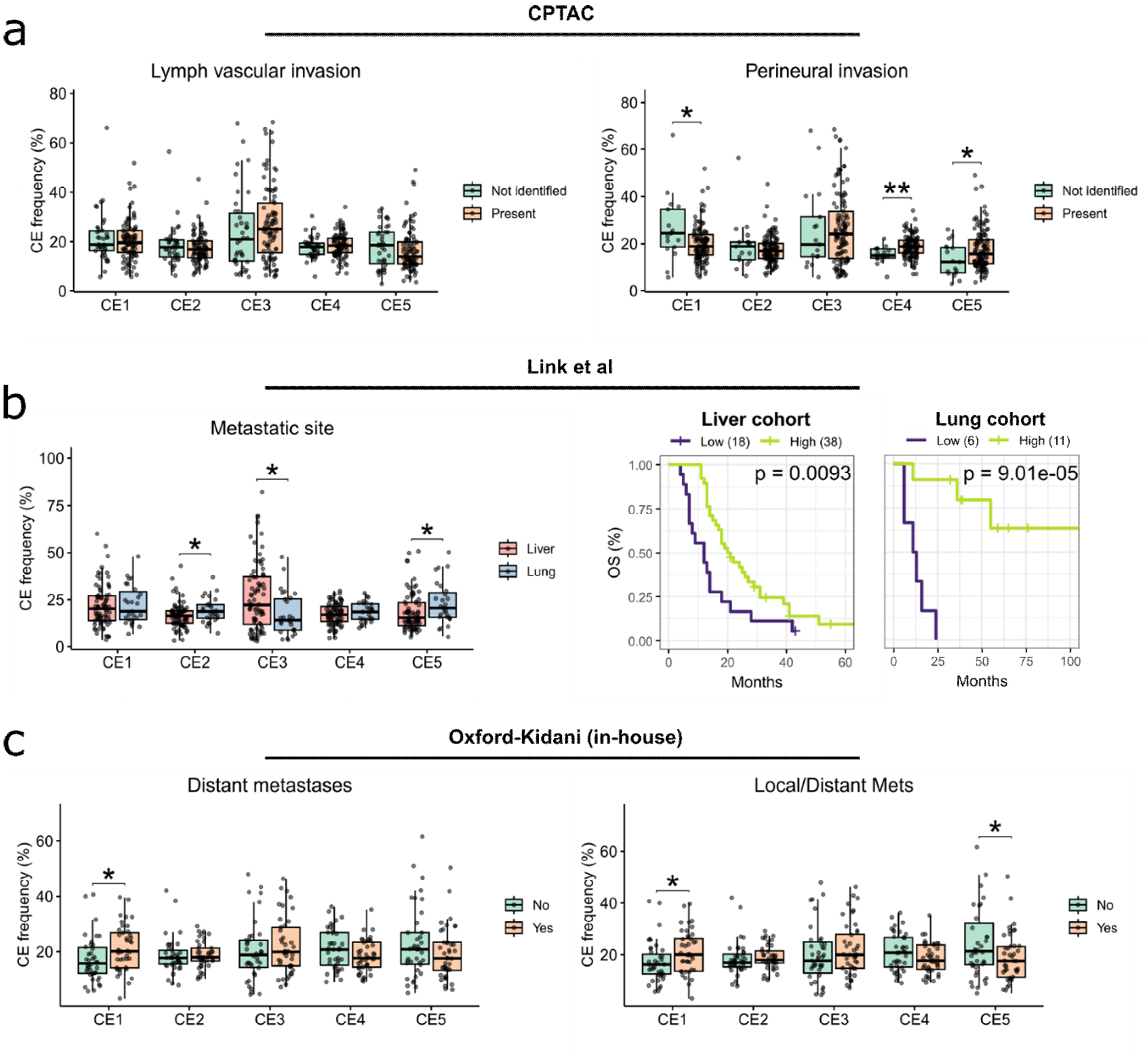


Figure 5.5 CEs and metastasis. (a) Box plots showing the comparison of CE abundance between presence or absence of lymph vascular invasion (left) and perineural invasion (right) from the CPTAC cohort. **(b)** Box plots showing the comparison of CE abundance between the lung and liver cohorts from the Link et al study. Kaplan-Meier survival curves for CE5 level abundances in both the liver (left) and lung cohorts (right). **(c)** Box plots showing the comparison of CE abundance and evidence of distant metastasis (left) and both local and distant metastases (right) from the Oxford-Kidani cohort. **(a-c)** Statistical significance was calculated using a two-sided Wilcoxon rank-sum test. *P<0.05, **P<0.01, ***P<0.001.

Next, we compared our CET classification with TME-based prognostic subtypes from previous studies[116] [203] [204] [41] (Fig. 5.6a). Results show that our CET subtypes displayed poor to moderate overlap with previous classifications, highlighting the potential of our CET classification to offer new insights into the PDAC TME. We then investigated the prognostic value of our CET subtypes by adjusting for the basal-like and classical subtypes. We first stratified patient samples into the consensus basal-like, mixed, and classical subtypes through unsupervised clustering of the Moffitt et al[40], Collisson et al[39], and Bailey et al[41] gene signatures (Fig. 5.6b). After excluding the mixed group, we revealed that the CET subtypes were independently prognostic in both the classical and basal-like subtypes, with CET3 and CET5 displaying opposing survival outcomes (Fig. 5.6c). The CETs were also significantly prognostic in multivariate Cox regression for both OS and DFS (Fig. 5.6d).

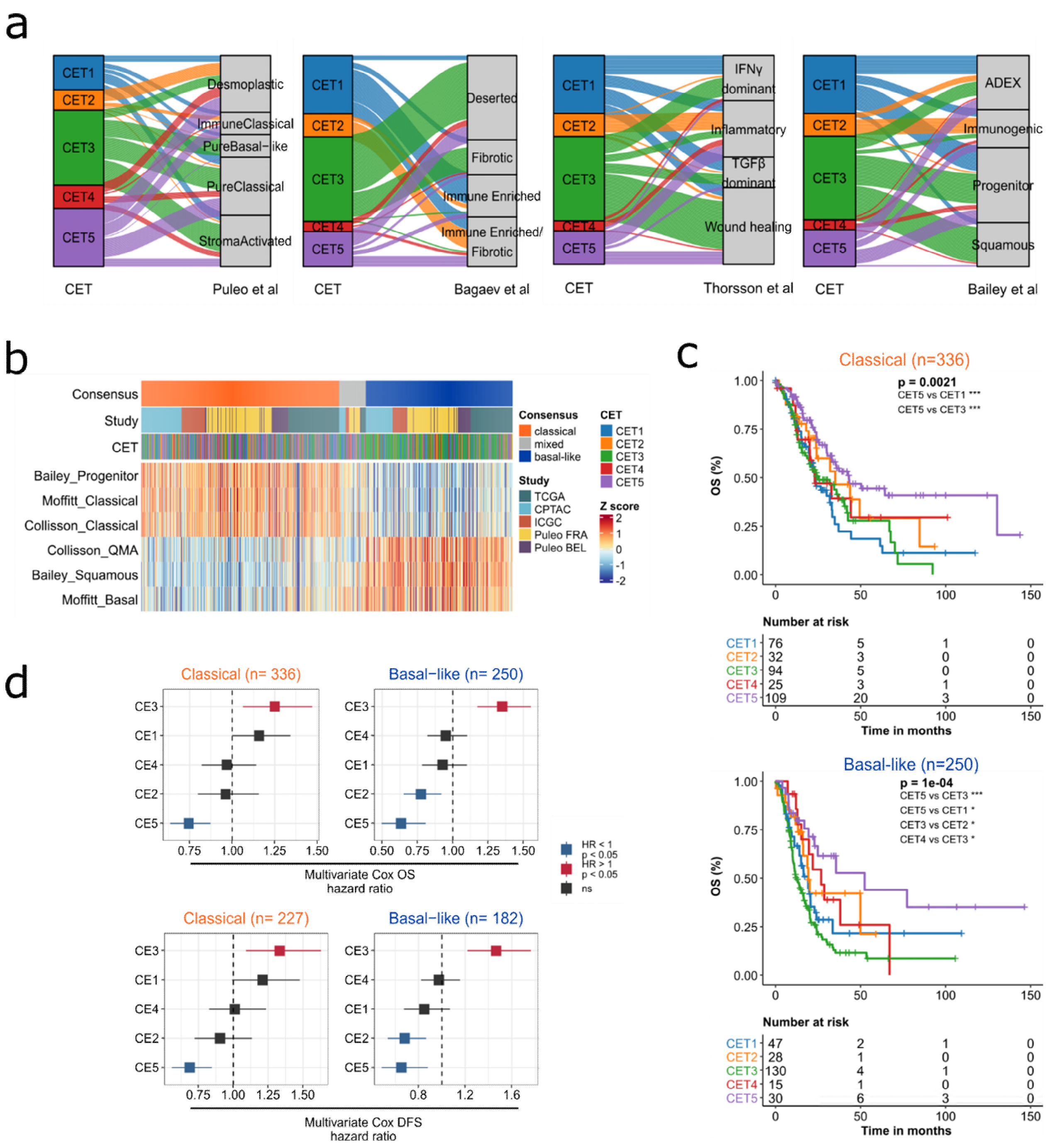


Figure 5.6 Comparison of the CET classification system with previous subtypes. (a) Alluvial plot showing the overlap between the stratified CETs and previous TME-based classification from Puleo et al, Bagaev et al, and Bailey et al. **(b)** Heatmap showing the two-consensus classical and basal-like subtypes across all five PDAC cohorts. Patient samples with enriched of both classical and basal-like signatures were classified as "mixed". **(c)** Kaplan-Meier survival curves for the assigned CETs using all PDAC cohorts for classical (top) and basal-like tumours (bottom). Statistical significance was determined using the

two-sided log-rank test for each CET pair. **(d)** Forest plots showing the association of CE abundances as a continuous variable with overall survival using multivariate Cox proportional hazards regression.

5.5 Association of CEs with treatment responses

Recent studies have shown the effect of neo-CTx in modulating the TME, and potentially priming the stroma to be more susceptible to immunotherapeutic interventions[205]. However, the effect of neo-CTx in PDAC cellular communities remains poorly understood. Therefore, we investigated the association of CE abundance following neo-CTx (FFX) from Zhou et al[49] (n = 97, neo-CTx = 33, naïve = 64), and compared results with the single-cell atlas. CE1 ($P = 6.1\times10^{-5}$) abundance was significantly depleted following neo-CTx, whilst CE4 ($P = 0.02$) was significantly enriched, consistent with the single-cell atlas (Fig. 5.7a). However, unlike the single-cell atlas, CE2 was also significantly enriched ($P = 0.002$). In the enriched group following neo-CTx, the single-cell atlas and bulk dataset shared many cell states (41.2%, 14 out of 34), including activated lymphoid and myeloid cells, as well as neurotropic fibroblasts (Fig. 5.7b-c). In the depleted group, 22.2% (4 out of 18) of the cell states were shared between the two datasets. Both M2-like macrophage populations (Mϕ_c2_CCL18 and Mϕ_c4_VEGFA) were depleted following neo-CTx, along with naïve T cells and pericytes (Fig. 5.7b-c). Notably, all proliferating immune cells from CE1 were depleted in the bulk dataset, suggesting that the systemic FFX chemotherapy may cause off-target effects (Fig. 5.7b). Moreover, the TLS-like CE2 was significantly associated with improved survival (HR = 0.30, $P = 0.01$), indicating enhanced anti-tumour immunity, which is consistent with recent findings in hepatocellular carcinoma[206]. In contrast, enriched CE3 abundance was associated with worse outcomes (HR = 1.5, $P = 0.1$).

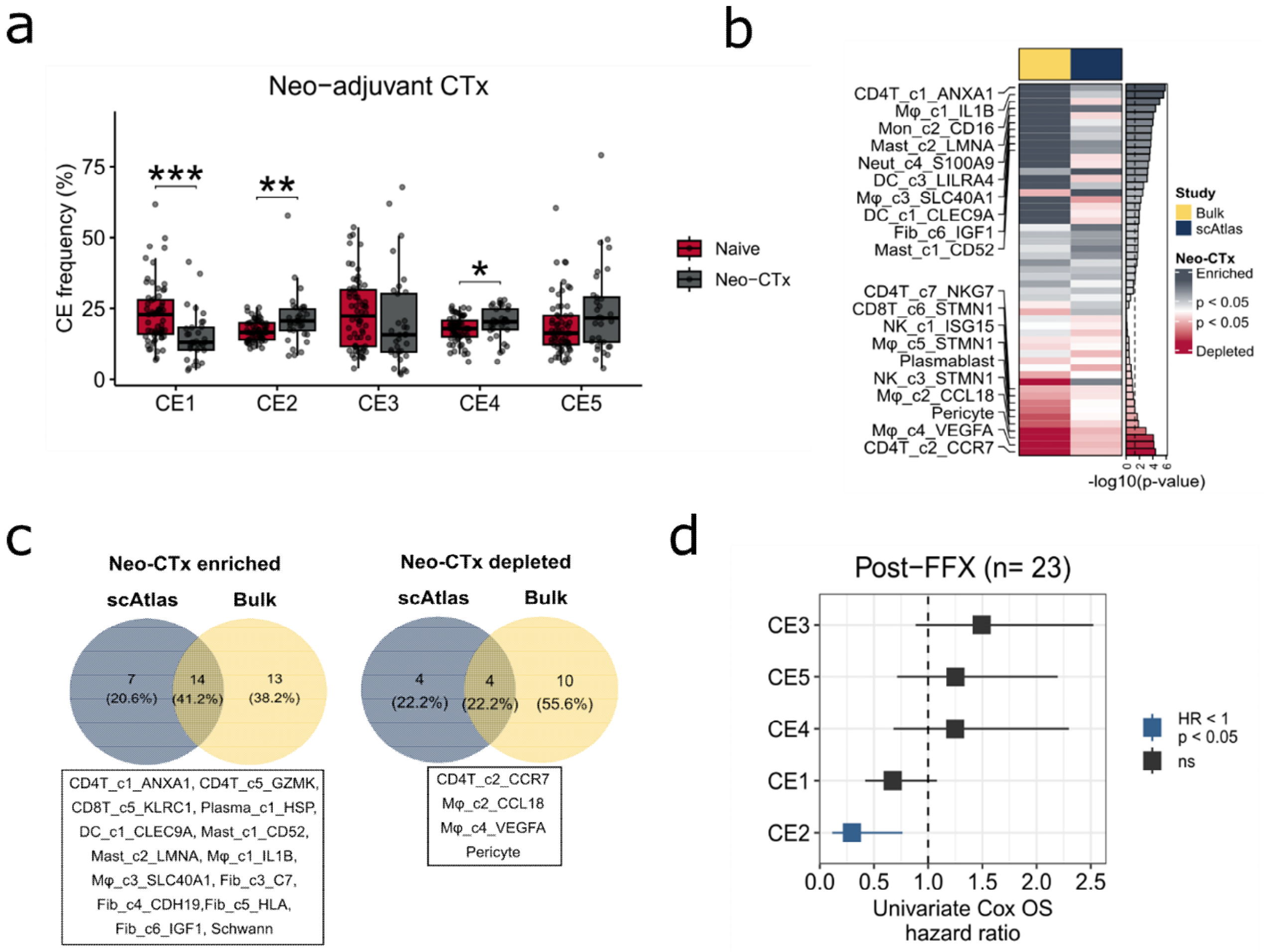


Figure 5.7 CE dynamics following neo-adjuvant chemotherapy. (a) Box plot comparing CE abundances between treatment-naïve and Neo-CTx treated samples from the Zhou et al study. Statistical significance was calculated using a two-sided Wilcoxon rank-sum test. **P*<0.05, ***P*<0.01, ****P*<0.001. **(b)** Concordance heatmap showing the enrichment or depletion of TME cell states following Neo-CTx in both the bulk and scAtlas cohorts. Bar plot on the right represents the mean of significance association for each cell state. **(c)** Venn diagrams showing the overlapping cell states enriched (left) and depleted (right) following neo-CTx. **(d)** Forest plots showing the association of CE abundances as a continuous variable with overall survival using univariate Cox proportional hazards regression.

Next, we utilized our in-house Oxford-Kidani and TCGA cohorts to investigate whether CE abundances could predict responses to adjuvant GemCap, which is a current standard-of-care treatment for PDAC patients[207]. We used DFS as a clinical endpoint. Results show that in both Oxford-Kidani and TCGA, CET1 was significantly associated with worse DFS compared to CET5 (Fig. 5.8a). We then investigated the association of cell states with GemCap responses (Fig. 5.8b-c). Both studies displayed strong concordance between the estimated cell state abundances and survival (R = 0.32, *P* = 0.018, Fig. 5.8b-c) Results show that the proliferating macrophage population (Mφ_c5_STMN1) demonstrated the worst outcomes, which is consistent with a recent multi-omics study in PDAC[208]. This was followed by the CE3-specific fibroblasts (Fib_c1_MMP11 and Fib_c2_CXCL8), linking fibrosis with chemoresistance. In contrast, the neurotropic Fib_c6_IGF1 population was associated with most favourable outcomes, followed by cDC1 (DC_c1_CLEC9A) and multiple endothelial cell states (Fig. 5.8d-e). Taken together, these findings demonstrate clear associations between CE abundances and chemotherapeutic responses, suggesting their potential utility as biomarkers to guide treatment selection. For example, identifying CE5-enriched patients who may benefit most from GemCap chemotherapy.

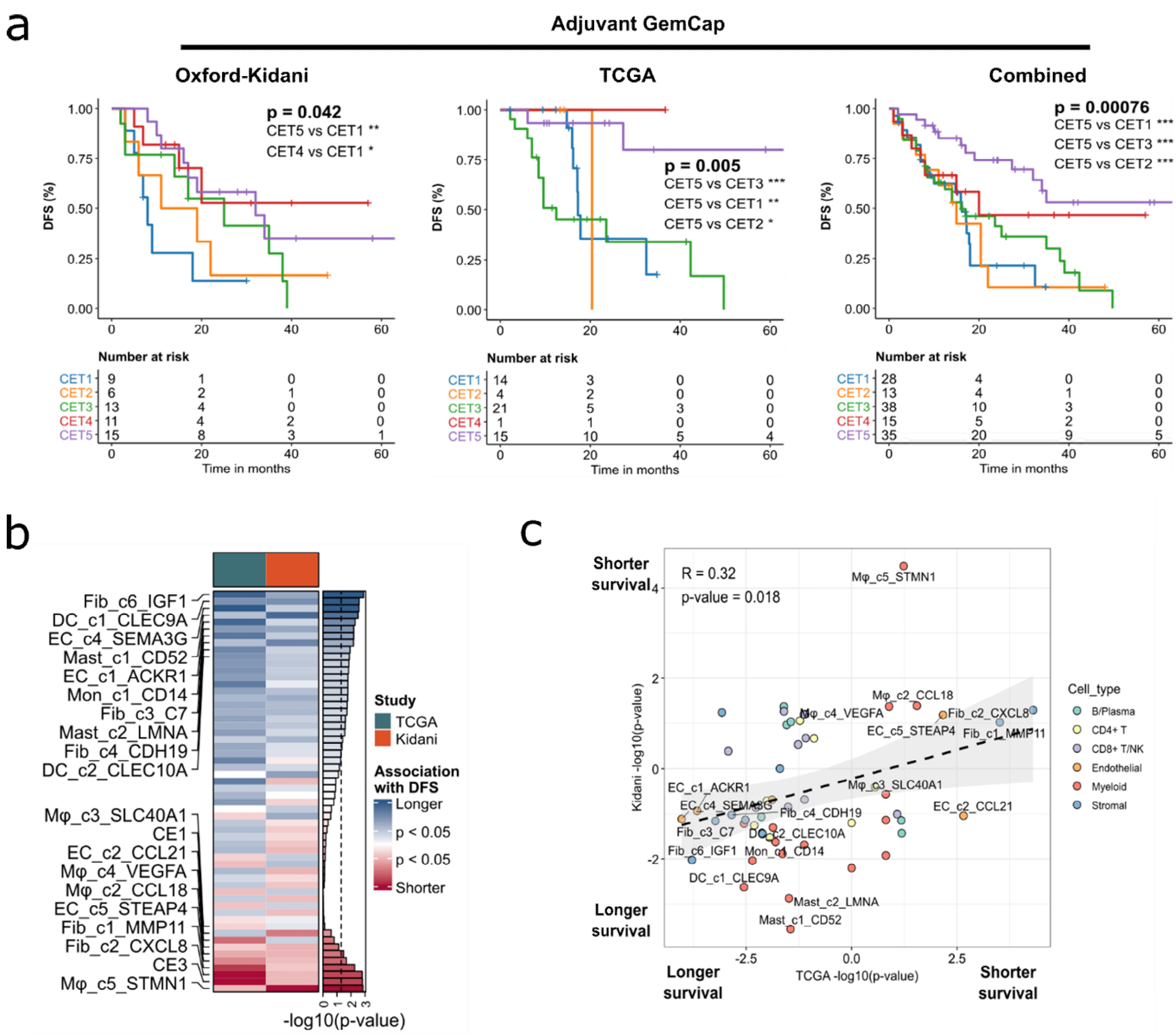


Figure 5.8 Association of CEs with adjuvant chemotherapy responses. (a) Kaplan-Meier DFS survival curves for the assigned CETs from the in-house Oxford-Kidani cohort (left), TCGA (middle), and combined (right) in patients treated with adjuvant GemCap. Statistical significance was determined using the two-sided log-rank test for each CET pair. **(c)** Kaplan-Meier OS (left) and DFS (right) survival curves for the assigned CETs from the TCGA cohort in patients treated with adjuvant GemCap. Statistical significance was determined using the two-sided log-rank test for each CET pair. The top p-value represents the four-way log-rank result. **(d)** Concordance heatmap showing the association TME cell states with disease recurrence in patients treated with adjuvant GemCap. Bar plot on the right represents the mean of significance association for each cell state. **(e)** Correlation scatter plot comparing the association of TME

cell states with disease recurrence between both cohorts. Points are coloured based on cell type. Statistical significance was calculated using Pearson's correlation.

Given the strong association of both M2-like macrophages and fibrotic stromal cells with adjuvant chemoresistance, we explored TME-based immunohistochemistry (IHC) scores of our in-house cohort to identify surrogate biomarkers for these CEs. We selected CD68 and αSMA as IHC markers, as CD68 serves as a surrogate for macrophage abundance, whereas αSMA marks activated fibroblasts and stromal fibrosis. Results show that dual α-SMA+CD68 low scoring was associated with the chemo-responsive CE5 while a dual positive staining was associated with the fibrotic CE3 (Fig. 5.9a-b). Moreover, the CD68$^{\text{low}}$ αSMA$^{\text{high}}$ was associated with worse response compared to CD68$^{\text{low}}$ αSMA$^{\text{low}}$ ($P = 0.015$) (Fig. 5.9c). Although, there was no significant difference in CD68$^{\text{high}}$ αSMA$^{\text{high}}$ compared to CD68$^{\text{low}}$ αSMA$^{\text{low}}$, suggesting that stromal fibrosis is more strongly linked to chemoresistance.

However, when using CE1-defined macrophage abundance instead of general CD68 scoring, the CE1-Mϕ$^{\text{high}}$ αSMA$^{\text{high}}$ group demonstrated worst survival, with the shortest median DFS (10 months), compared with the CE1-Mϕ$^{\text{low}}$ αSMA$^{\text{high}}$ and CE1-Mϕ$^{\text{low}}$ αSMA$^{\text{low}}$ groups (Fig. 5.9d). These findings suggest that more precise biomarkers, such CE1-specific macrophage markers are warranted to better characterize macrophage abundance and improve patient stratification for predicting chemotherapy responses.

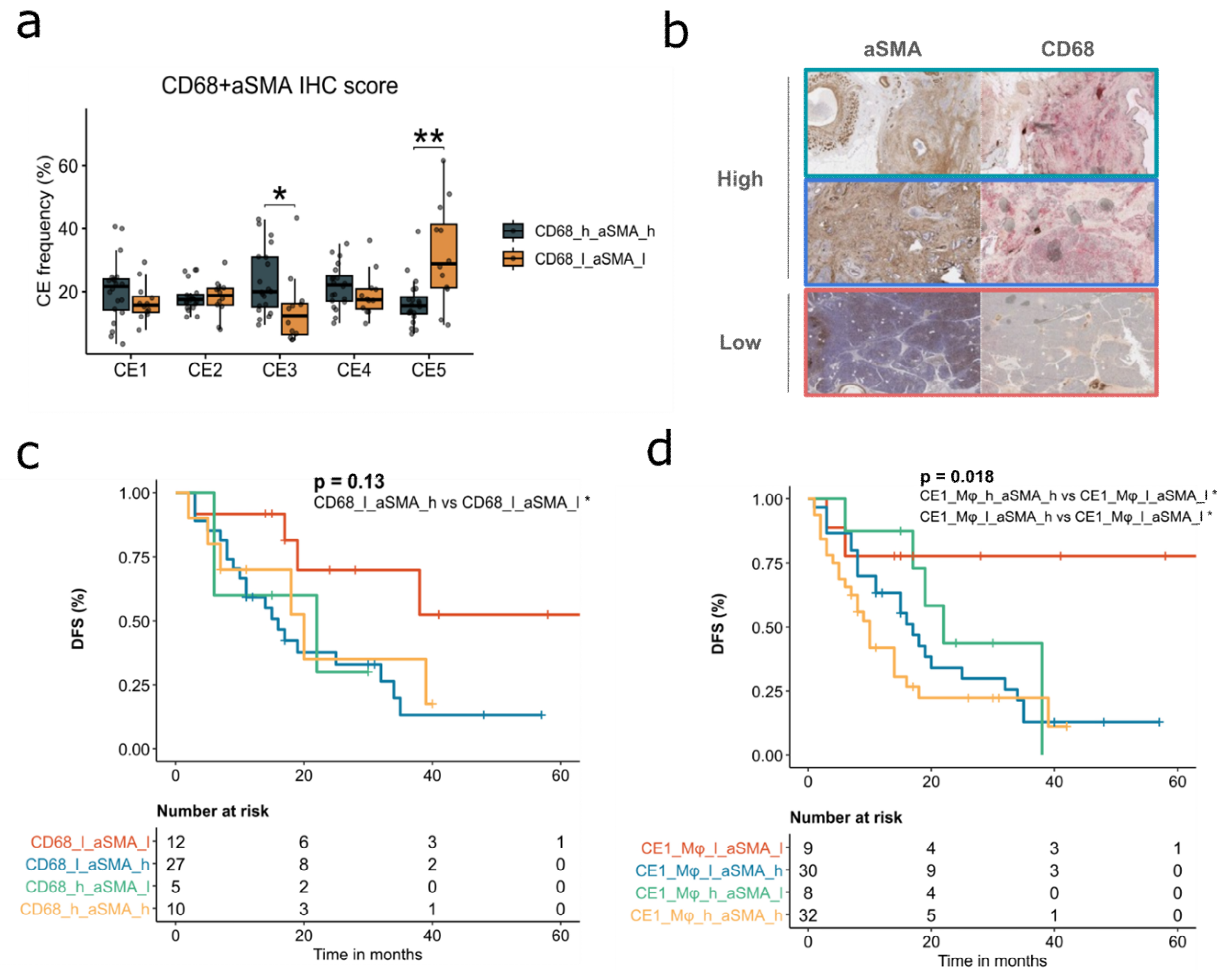


Figure 5.9 Surrogate IHC biomarkers for CEs. (a) Box plots comparing CE abundance between the dual CD68^{high}αSMA^{high} and the CD68^{low}αSMA^{low} groups. Statistical significance was calculated using a two-sided Wilcoxon rank-sum test. *$P<0.05$, **$P<0.01$, ***$P<0.001$. **(b)** Representative IHC staining for both CD68 and αSMA stains in two CD68^{high}αSMA^{high} and one CD68^{low}αSMA^{low} patient group. **(c)** Kaplan-Meier DFS curves for the staining combinations in patients treated with adjuvant GemCap. Statistical significance was determined using the two-sided log-rank test for each group pair. **(d)** Kaplan-Meier DFS curves for the CE1 Macrophage and αSMA group combinations in patients treated with adjuvant GemCap. Statistical significance was determined using the two-sided log-rank test for each group pair.

Cancer immunotherapeutic approaches, including ICB therapies, have shown significant efficacy in various cancer types, however, their success in PDAC has been limited[209]. Therefore, we investigated whether the CE classification system can be used to predict response to ICB. For this analysis, we leveraged 7 publicly available bulk RNA-seq datasets from diverse cancer cohorts, that were annotated with RECIST-based response evaluations and survival outcomes following treatment with anti-PD1, anti-PD-L1, anti-CTLA4, or combined anti-PD1 and anti-CTLA4 therapies.

Results show that the immunosuppressive/proliferative CE1 demonstrated the best response to immunotherapy compared to other CEs and cell states (Fig. 5.10a). In contrast, the fibrotic CE3 demonstrated the worst response, consistent with the immune-exclusion phenotype of this CE. We also compared our CE classification with published ICB-response gene signatures, including the ‘T cell inflamed GEP’, ‘IFNG’, and ‘TLS’, as well as ICB-response marker genes, including *PDCD1, CD274*, and *CTLA4*. Notably, CE1 abundance outperformed all other indicators, including tumour mutational burden (TMB), highlighting the strong predictive value of this CE. Specifically, both CD8T_c6_STMN1, and NK_c3_STMN1 cell states, both belonging to CE1, were associated with best ICB response. Conversely, the VEGFA$^+$ macrophage population (Mϕ_c4_VEGFA) showed worse responses to immunotherapy, in line with previous findings[188]. Additionally, the average abundance of both the CE1 cell states (CD8T_c6_STMN1, and NK_c3_STMN1) demonstrated superior predictive power to various ICB therapies, including anti-CTLA4 (AUC = 0.82), anti-PD1 (AUC = 0.79), and anti-PD-L1 (AUC = 0.8), compared to the public ‘TLS’, and ‘CD8’ markers (Fig. 5.10b).

Immunotherapies have been shown to reshape the TME[210], thus tracking these changes can provide critical insights into ICB response. To investigate this, we analysed CE frequencies in bulk RNA-seq data from a cohort of eight PDAC patients treated with anti-PD1 and radiation therapy (RT). This included patient samples collected at baseline (pre-ICB), after ICB treatment (post-ICB), and following RT (post-RT)[119]. Results reveal substantial changes in CET assignments along treatment stages (Fig. 5.11a). The single patient that displayed complete response (CR) developed a CET1 post-RT, further validating the association of CE1 with effective ICB response. Additionally, the patient that displayed partial response (PR) developed CET1 post ICB, followed by the fibrotic CET3 post-RT. This suggests that development of the fibrotic/immune-exclusive TME prevents ICB efficacy despite the presence of CE1. This reduction in efficacy is also seen when calculating a CE1:CE3 ratio (Fig. 5.11a). However, this analysis was limited to primary tumours, therefore, we could not assess the development of CET3 or CE1:CE3 ratio dynamics in metastatic tissues.

To further validate these findings, we analysed the Riaz et al melanoma cohort which included pre-ICB and post-ICB biopsies collected from patients treated with anti-PD1 immunotherapy[119]. Similar to the PDAC cohort, the majority of CR patients developed a CET1 phenotype post-ICB without a pre-existing CET3 (Fig. 5.11b). Additionally, a high CE1:CE3 ratio was significantly associated with response when comparing CR vs PD (*P* = 0.01), and PR vs PD (*P* = 0.04), only during the post-ICB group. Collectively, these findings highlight the importance of tracking TME dynamics during ICB immunotherapy, with patients developing CE1 from non-CE3 tumours respond best to immunotherapy.

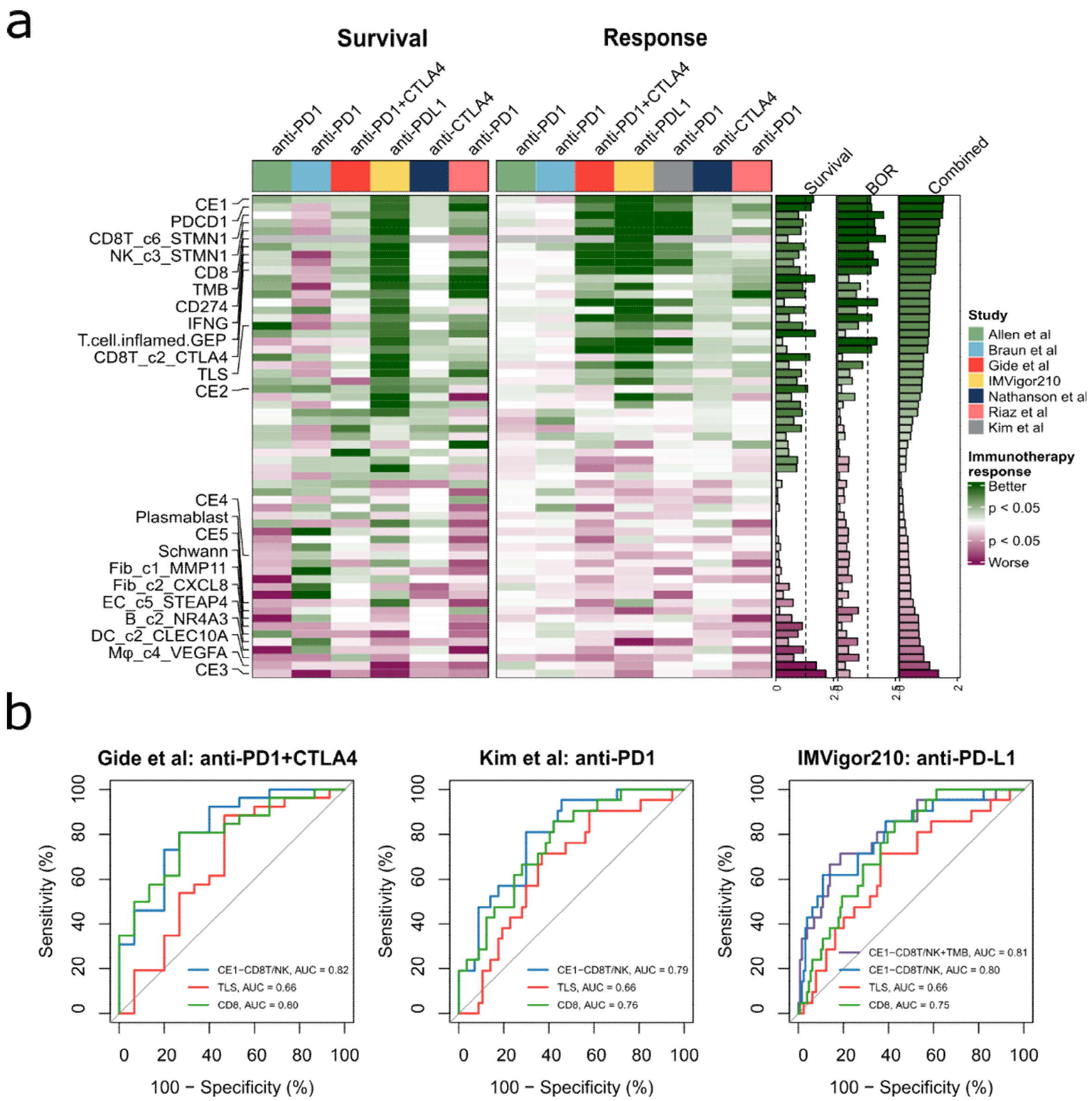


Figure 5.10 Association of CEs with ICB responses. (a) Concordance heatmap showing the association of CE, cell state and public biomarker levels with ICB response across immunotherapy cohorts. Survival bar plots on the right indicate the average significance association (-log10(p-value)) for each variable with overall survival (left heatmap), while the BOR bar plots indicate the average significance association (-log10(p-value)) for each variable with RECIST response (right heatmap). An average of both survival and BOR bar plots is indicated as a combined performance score and is used to order all variables from best to worse ICB response. **(b)** ROC curves comparing the predictive performance of the CE1 CD8T/NK cells with other ICB-response markers across immunotherapy cohorts. AUC: Area under the curve.

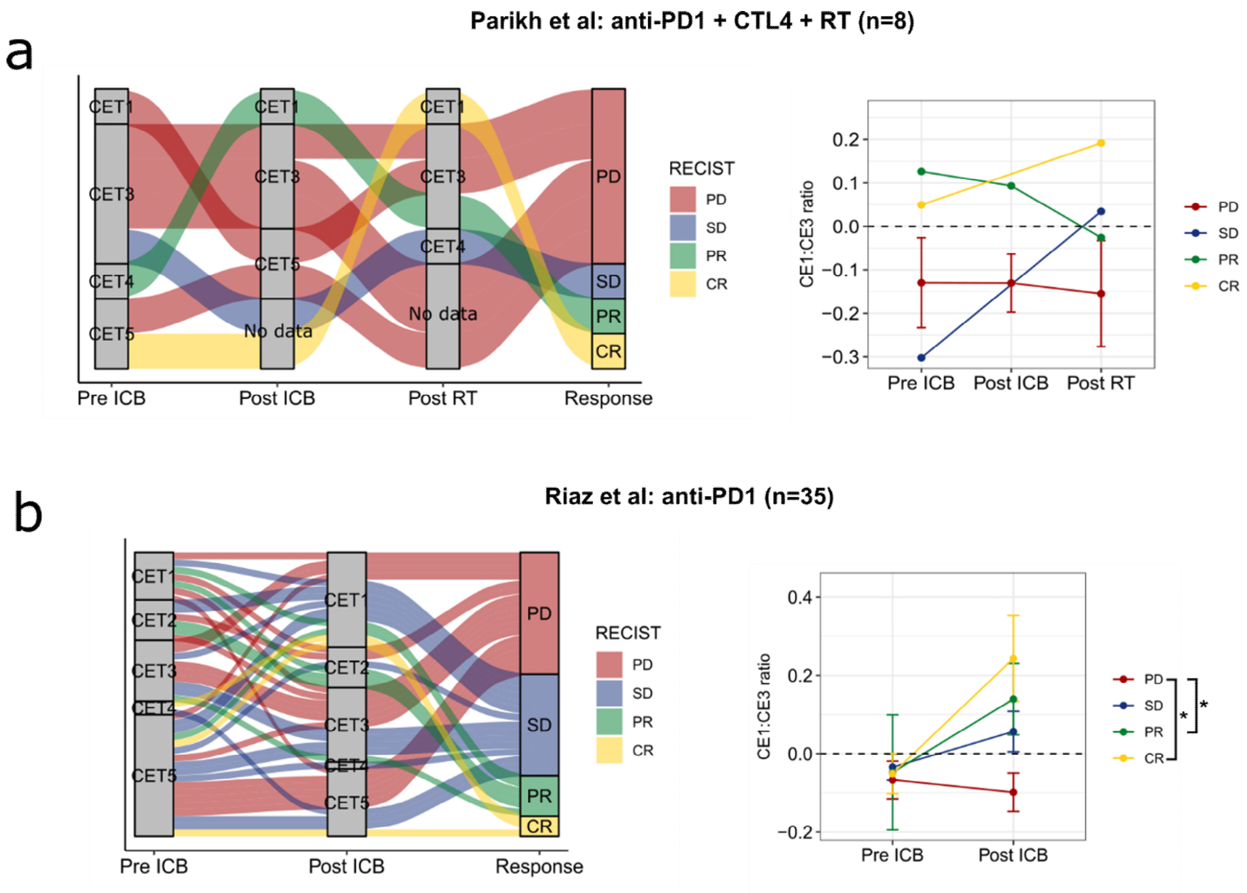


Figure 5.11 TME dynamic changes along ICB treatment stages (a) Alluvial plot showing the patient stratification to CETs in the pre-ICB, post-ICB, and post-RT groups from the Parikh et al PDAC cohort (left). Paired analysis of the CE1:CE3 ratio levels across pre-ICB, post-ICB, and post-RT groups (right). **(b)** Alluvial plot showing the patient stratification to CETs in the pre-ICB, and post-ICB groups from the Riaz et al melanoma cohort (left). Paired analysis of the CE1:CE3 ratio levels across pre-ICB, and post-ICB groups (right). Statistical significance within the on-treatment response groups was determined using an unpaired Wilcoxon rank-sum test. **P*<0.05, ***P*<0.01, ****P*<0.001. RT: Radiation therapy; PD: Progressive disease; SD: Stable disease; PR: Partial response; CR: Complete response.

Cytokines are key immune response regulators, and hold promise as biomarkers for predicting patient outcomes and treatment responses[211]. However, current pre-clinical models, including patient-derived cell lines, xenografts (PDX), and organoids, do not capture the TME complexity, and therefore, cannot provide TME-derived cytokines needed for effective patient stratification[212]. As a result, the O'Neill group has established a personalized organotypic "avatar" system from resected tissue that successfully maintains the TME architecture and predicts clinical responses[213]. Therefore, we utilized this *ex vivo* system to characterize CE-based cytokine profiles in conditioned media from five patients. We first performed differential expression of the down-sampled atlas and filtered enriched cytokines that were included in the cytokine profile array (60 cytokines) (Fig. 5.12a). CE2 did not display unique cytokines due to the limited cytokine panel that lacked CE2-specific cytokines including TNF-α, and IL-16.

Interestingly, the patient that displayed favourable response to stereotactic body radiotherapy (SBRT) (Patient_1), exhibited high levels of CE5 cytokines and lacked CE3-specific cytokines (Fig. 5.12b-d). Conversely, the patient that displayed metastasis (Patient_3), showed elevated levels of CE3 cytokines. Patients 4 and 5 displayed the highest cytokine levels of CE3 along with other CEs, however no clinical follow up data was obtained for both patients. In summary, these findings provide a proof of concept for the potential utility of integrating CE-based biomarkers with the *ex vivo* avatar model system to effectively predict patient prognosis and treatment responses.

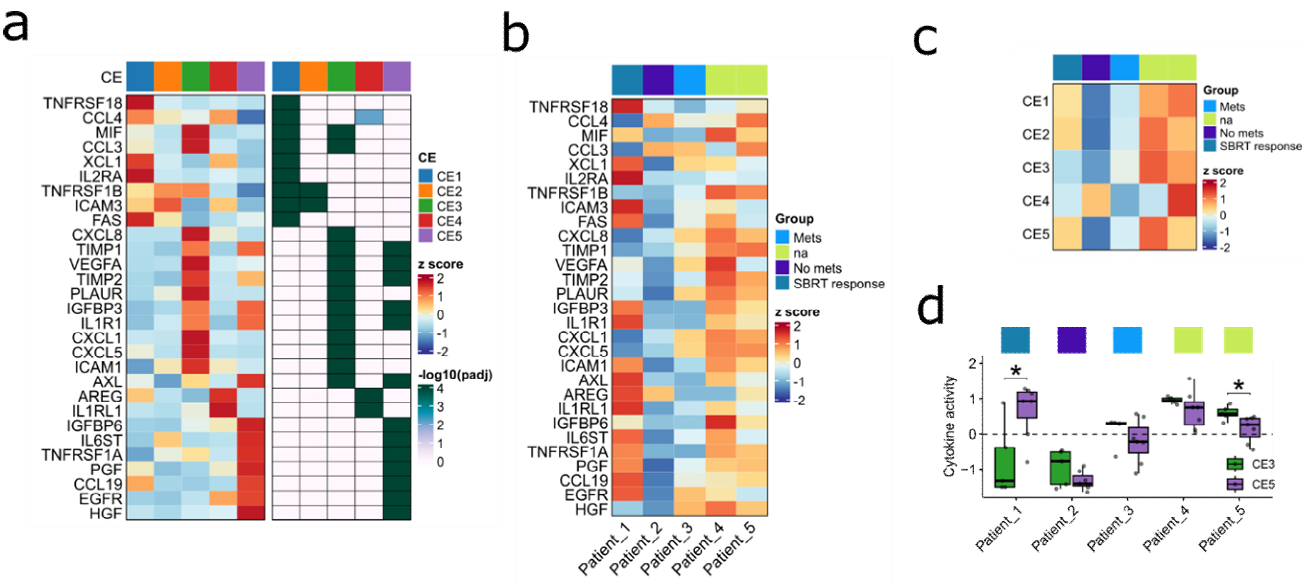


Figure 5.12 CE-based cytokine profiles in patient-derived avatars. (a) Heatmap showing the distinct CE-specific cytokines identified using differential gene expression analysis. Left heatmap shows the scaled average expression while the right heatmap shows the binary significance across CEs (padj < 0.05). **(b)** Heatmap showing the CE-based cytokine protein levels across patient samples. The normalized protein levels were scaled. **(c)** Heatmap showing the average levels of CE-based cytokines across patient samples (columns). **(d)** Box plots comparing CE3 and CE5 cytokine activity across patient samples. Each point indicates a cytokine belonging to the corresponding CE. Statistical significance was determined using an unpaired Wilcoxon rank-sum test. **P*<0.05, ***P*<0.01, ****P*<0.001.

5.6 Conclusions and Discussion

Due to the frequently late diagnosis, early metastasis, and high therapeutic resistance, PDAC prognosis remains poor[191]. Our findings reveal previously uncharacterized CEs that could provide clinical utility for patient prognosis and treatment selection. Furthermore, given the limited therapeutic options, early detection of PDAC through diagnostic biomarkers is essential for improving survival. Although our analysis was not primarily focused on early detection, the CE5 phenotype may still provide key diagnostic insights, as this cellular ecosystem is particularly enriched normal tissue but also present in PDAC tumours. A key characteristic of this ecosystem is the complement-enriched fibroblast population (Fib_c6_IGF1) expressing high *IGF1* levels and other markers. Recent reports have shown that IGF-1 is connected with early carcinogenesis in animals and humans[214]. However, studies have shown that IGF-1 and IGFBP-2 levels did not perform better than CA19-9 in differentiating between PDAC from chronic pancreatitis (CP)[215]. Other CE5-specific cytokines could help diagnose PDAC in CP patients. For instance, *PDGFB* which we found enriched in CE5 through our ligand-receptor pair analysis, was recently reported as a candidate biomarker enriched in PDAC compared to CP[216]. Therefore, screening CE-specific cytokines may uncover additional diagnostic biomarkers for the early detection of PDAC, potentially enhancing the sensitivity of the established marker CA19-9.

The FOLFIRINOX chemotherapeutic regimen is considered the most effective first-line treatment for patients with locally advanced or metastatic PDAC, displaying prolonged survival compared to gemcitabine treatment[217]. Despite the improved response rates, around 25% of patients develop disease progression during treatment[218]. Our results in both the single-cell atlas and bulk validation datasets show that FOLFIRINOX significantly remodels the TME with the reduction in immunosuppressive/proliferative CE1 an increase in activated immune cells (CE4) and neurotropic fibroblasts (CE4 and CE5). This similar TME remodelling has been associated with improved survival following neo-CTx in both PDAC and CRC tumours[150 160]. However, re-emergence of the immunosuppressive CE1 has been linked to progressive disease in CRC[160]. Our results show that high CE1 abundance is associated with presentation of early distant metastases at diagnosis. This suggests that a subset of patients who initially present with CE1 may respond well to FFX (through reduction of CE1 immunosuppressive cells) but may later relapse, with CE1 re-establishing in metastatic sites. Therefore, further investigation into the molecular mechanisms driving metastasis in CE1 tumours is warranted.

Moreover, CE1 abundance shows strong predictive value for adjuvant GemCap resistance. We found that the proliferating macrophage population (Mϕ_c5_STMN1) was more strongly associated with resistance than any other TME cell states. This aligns with findings from Zhang et al[208], where they showed that proliferating macrophages survive chemotherapy by reducing gemcitabine absorption via enhanced deoxycytidine (dC) production as well as promoting fibrosis. Conversely, we also reveal that CE5 tumours are strongly associated with improved survival following GemCap. Therefore, we propose that identifying CE5 PDAC tumours, either through IHC scoring of both CD68 and α-SMA or secreted cytokine profiling, could help stratify GemCap responders from the resistant CE1/CE3 groups. Additionally, this CE classification system could help in minimizing chemotherapy-induced toxicity, for example, patients with CE1 tumours may be directed to FOLFIRINOX, while CE5-enriched patients could benefit from the less toxic GemCap regimen[14].

Despite successes of ICB in other solid cancers, PDAC has been largely refractory to ICB[219]. Our results show that CE1 abundance outperformed other known biomarkers predicting ICB responses across cancer types, including TLS[220]. As a result, measuring CE1 abundance may act as a powerful approach for identifying PDAC patients who should be treated with ICB. Notably, CE1 did not display the classical immune-hot phenotype associated with favourable responses to ICB across cancer types[221]. Instead, CE1 consisted of immunosuppressive and proliferating M2 macrophages, proliferating CD8+ T and NK cells, and Tregs. Similarly, a recent study in sarcoma identified an ecotype characterized with M2-like macrophages that was resistant to chemotherapy but responsive to ICB[222]. However, our analyses using both the single-cell atlas and spatial transcriptomics revealed that CE1 and CE3 frequently co-exist within individual tumours, which may limit effective ICB responses in pancreatic cancer.

Therefore, the CE1 to CE3 ratio, could serve as a more accurate biomarker for predicting ICB efficacy, as supported by our findings in the longitudinal immunotherapy cohorts.

Moreover, a recent Phase 2 study combining nivolumab with modified FOLFIRINOX in metastatic PDAC failed to show meaningful increase in efficacy[223]. Our findings indicate that proliferating CD8+ T and NK cells within CE1, which were most strongly associated with ICB response, were depleted following FOLFIRINOX chemotherapy. Consequently, the loss of ICB-responsive cells by systemic chemotherapy could explain the limited efficacy of a-PD-L1 therapy. These results highlight the need for alternative treatment strategies to enhance immunotherapeutic efficacy in PDAC.

Chapter 6

Tumour cell heterogeneity and association with CEs

6.1 Introduction and Aims

In recent years, large-scale scRNA-seq and spatial transcriptomic studies have provided deeper insights into the ITH across diverse cancer types. For example, Gavish et al, identified 41 consensus meta-programmes (MPs) spanning 24 tumour types, encompassing cellular processes such as cell cycle, stress responses, and lineage-specific programmes[224]. Similarly, a recent study in cervical squamous cell carcinoma highlighted distinct interactions between cancer cell states and immune and stromal components of the TME[225]. Despite these advances, PDAC ITH and how it shapes the TME remains underexplored, and there is an urgent need to investigate this heterogeneity beyond the conventional basal-like and classical subtype framework.

Furthermore, genomic alterations arising from chromosomal instability have been implicated in shaping PDAC subtype identity, with mutant *KRAS* amplifications particularly enriched in the aggressive basal-like subtype[42]. However, the relationship between genomic alterations and tumour cell states beyond the basal-like and classical subtypes remain poorly understood. Therefore, in this chapter, we aim to identify and characterize consensus tumour cell states of PDAC and to explore their interactions with TME. The specific objectives of this chapter include:

- Distinguish malignant cells from non-malignant epithelial cells using inferred copy number alterations and malignant gene signatures.
- Identify and characterize malignant cell states that underlie ITH.
- Investigate the associations of malignant cell states with the identified CEs and their distinct intercellular communications networks.
- Explore genomic alterations driving these changes in ITH.

6.2 Identification of PDAC tumour cells

We first subclustered the epithelial compartment into six phenotypes, including epithelial_1, ductal_1, ductal_2, acinar, tuft, and liver-specific hepatocytes (Fig. 6.1a). Epithelial_1 resembled tumour cells expressing high levels of established PDAC genes (*KRT19, S100A6, MUC1*) and was absent in NAT tissue (Fig. 6.1b). The *MUC5B*+ ductal_2 cluster also expressed malignant PDAC genes as well as normal ductal genes to a lesser extent. This suggests that ductal_2 cluster is an intermediate phenotype between normal and malignant PDAC cells[107].

To accurately distinguish tumour cells from non-malignant epithelial and TME cells, we analysed aneuploid CNAs profiles across all epithelial cells, as CNAs are a hallmark feature in around 90% of solid tumours[226]. We applied CopyKAT[125], which has been shown to outperform other gene expression-based CNA inference tools in a recent benchmarking study[227]. Each sample was processed individually to avoid batch effects, using immune cells as references (Fig. 6.1c).

To evaluate the accuracy of the tumour cell classification, we compared the genome-wide CNA frequencies of CopyKAT-predicted tumour cells with TCGA whole-exome sequencing (WES) data. The results demonstrated strong concordance with TCGA WES for both copy number amplifications ($R = 0.66$, $P = 8.1\text{x}10^{-06}$) and deletions ($R = 0.8$, $P = 8.9\text{x}10^{-09}$) (Supplementary Fig.

5a-d). Predicted tumour cells exhibited frequent amplifications in chromosome arms 1q, 8q, 7p, 20q, and 3q, and deletions in 17p, 18q, 9p, 6, 8p, and 3p, consistent with alterations commonly reported in PDAC[228 115].

Furthermore, to avoid false negatives in tumour cells with low CNA burden, we additionally evaluated epithelial gene signatures. Specifically, we assessed malignant-associated (*MUC1, CDH1, KRT19, FXYD3, KRT17, KRT7, S100A4, KRT8, S100A6, TFF1, TFF3*) alongside normal epithelial (*AMY2A, AMY2B, PRSS1, CFTR, SLC4A4, BICC1*) gene signatures to more accurately distinguish tumour from non-malignant cells (Fig. 6.1d-e)[107]. Additionally, results show that MT tumour cells displayed significantly higher CNA burden compared to PT ($P = 0.01$), highlighting the association of CNAs with disease progression and metastasis[229].

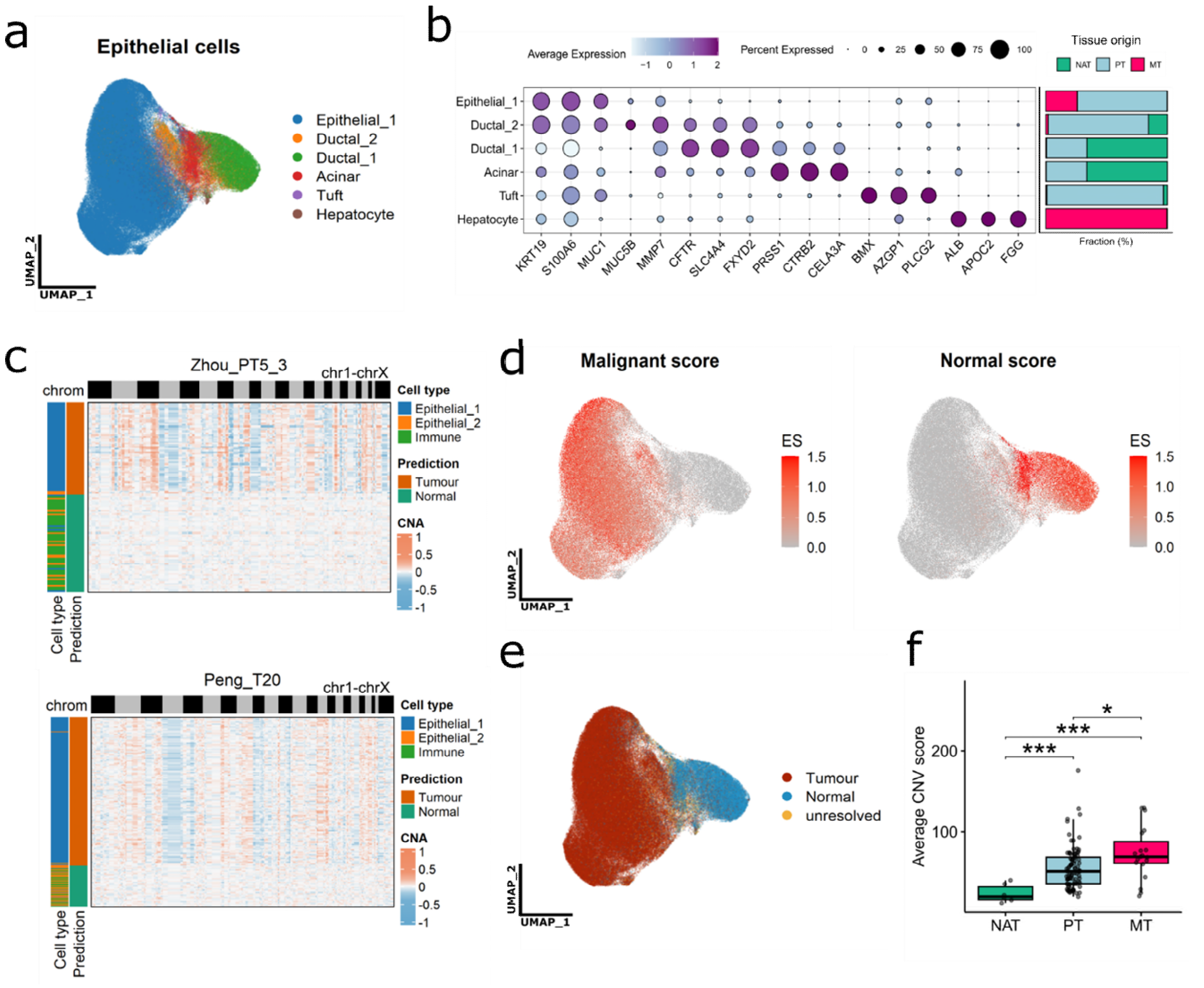


Figure 6.1 Identification of malignant PDAC cells. (a) UMAP embeddings of the epithelial cells coloured by the six identified clusters. **(b)** Dot plot showing the gene expression of markers specific to each epithelial cell cluster. Bar plot (right) showing the frequency of epithelial cell abundances in NAT, PT, and MT tissues. **(c)** CNA heatmaps of two individual samples estimated by CopyKAT. **(d)** UMAP embeddings of epithelial cells coloured based on the malignant (left) and normal epithelial (right) gene signature scores. **(e)** UMAP embeddings of epithelial cells coloured by the final predictions for distinguishing tumour from normal cells. **(f)** Box plots showing the differences in average CNA scores between tissue types. Statistical significance was determined using a two-sided Wilcoxon-rank sum test. *P<0.05, **P<0.01, ***P<0.001.

6.3 Identification and characterization of tumour ITH

To identify recurrent tumour cell states underlying intratumoural heterogeneity in PDAC, we analysed each sample containing more than 100 tumour cells separately. Briefly, clustering was performed at multiple resolutions and the top 50 significantly enriched genes from each cluster were defined as cluster-specific gene signatures. Across all samples, this yielded 621 unique gene signatures, which were subsequently subjected to consensus clustering (Methods 2.11.2). This analysis revealed six robust ITH clusters, determined based on the highest average silhouette width (Fig. 6.2a-b). Individual tumour cells were annotated based on the highest enrichment of the cancer cell state gene signatures (Fig. 6.2c, Supplementary Table 14).

Next, to characterize the cancer cell state phenotypes, we performed pathway enrichment analysis for Hallmark gene sets, the pan-cancer ITH MPs, and PDAC-specific gene signatures (Fig. 6.2d, Supplementary Table 11). State 1 was enriched with PDAC basal-like gene signatures (“Bailey Squamous” and “Moffitt Basal”), the “MP13 EMT” signature, and TGFβ signalling. This state highly expressed the basal-like *KRT17* (Fig. 6.2e)[48]. Conversely, state 2 was enriched with PDAC classical gene signatures and “MP38 Glutathione” gene signature. This state expressed high levels of trefoil-factor (*TFF1, TFF2*) and mucin (*MUC5AC, MUC13, MUC17*) genes, indicating a differentiated glandular phenotype (Fig. 6.2e). State 3 was enriched proliferation pathways including “MP1 Cell cycle G2/M”, “MP20 MYC”, and PI3K/AKT/mTOR signalling. Moreover, state 4 represented a progenitor phenotype, as it was enriched for the “Neural-like progenitor” and “Acinar-like” gene signatures, and both “Interferon/MHC-II” MPs (MP17 and MP18). This state also distinctly expressed the *OLFM4* intestinal progenitor marker (Fig. 6.2e)[230]. State 5 was enriched with “TNFα Signalling via NFκB”, “MP6 Hypoxia”, and “MP5 Stress” gene signatures.

Similar to state 3, state 6 cells were enriched with cell cycle gene signatures, primarily "MP2 Cell cycle G1/S", and "MP4 Chromatin" gene signatures.

We then correlated cancer cell state abundances with clinical features. Results show that both S3_Cycling_G2/M (*P* = 1.11×10^{-5}) and S6_Cycling_G1/S (*P* = 1.55×10^{-7}) cycling cell states were distinctly enriched in MT tissues (Fig. 6.2f). Conversely, the S4_Progenitor (*P* = 1.99×10^{-9}) was the most enriched in PT tissues compared to MT, followed by the S2_Classical (*P* = 0.007), and S1_Basal to a lesser extent (*P* = 0.03). Furthermore, only the S4_Progenitor was significantly enriched following neo-CTx (*P* = 0.009), consistent with the enrichment of neural-like progenitor (NRP) malignant cells following chemotherapy and radiotherapy reported by Hwang et al[61]. Lastly, both cell cycle states (S3 and S6) significantly co-enriched together (*R* = 0.69, *P* = 1.28×10^{-17}) while the majority of other cell states were uniquely enriched (Fig. 6.2g).

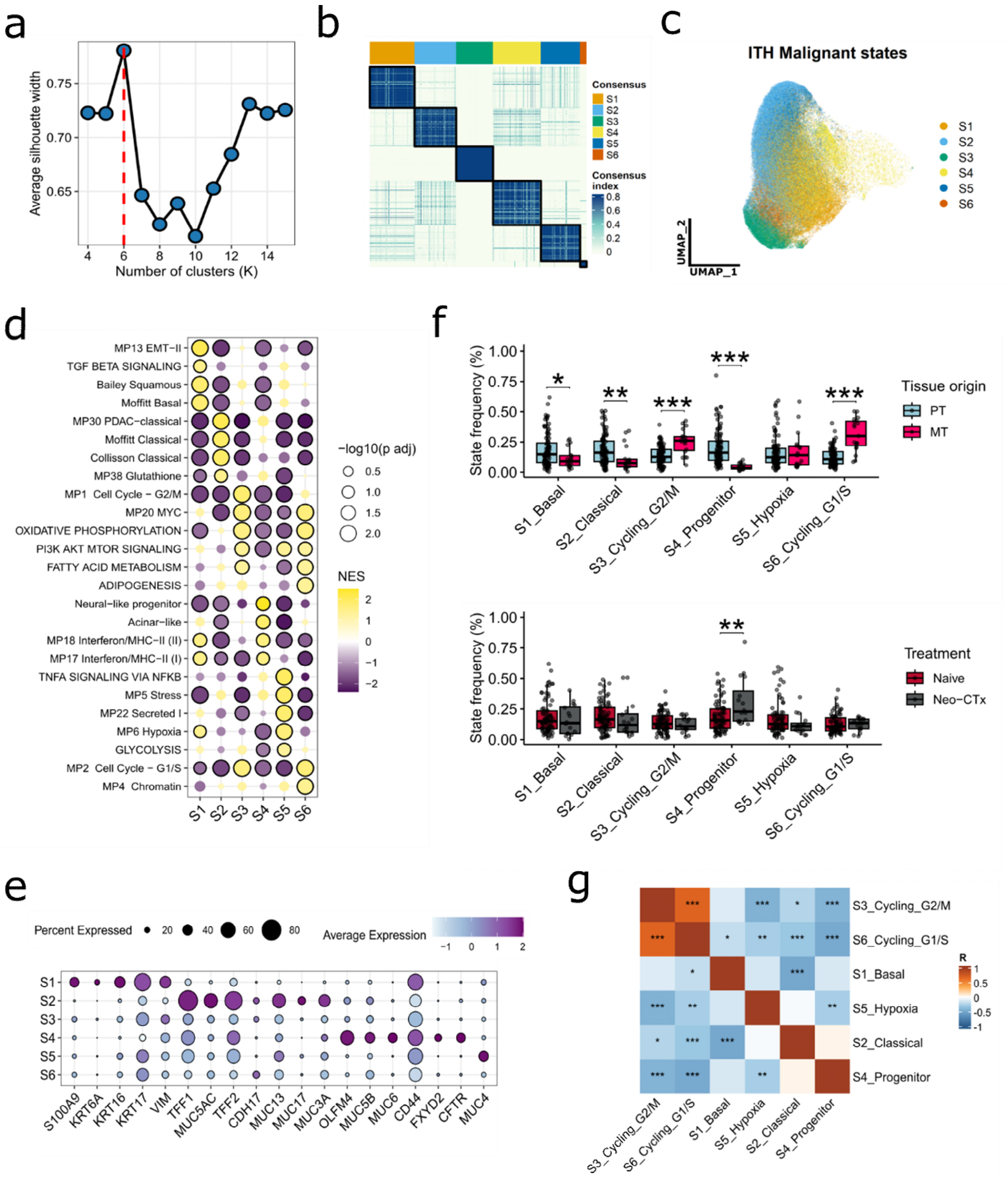


Figure 6.2 Identification and characterization of PDAC cell states. (a) Optimal consensus cluster selection based on the average silhouette width (ASW), with K = 6 identified as the optimal clustering number. **(b)** Heatmap of the consensus matrix, showing the five distinct consensus clusters. **(c)** UMAP embeddings of tumour cells coloured by cell states. **(d)** Gene set enrichment analysis between CEs. Colour of dots represent the normalized enrichment score (NES) and the size represents the -log10 (p-

value). Outlined dots represent significant differences (padj < 0.25). **(e)** Dot plot showing lineage specific markers across PDAC cell states. **(f)** Comparison of cell state frequency across tissue types and treatment conditions. Statistical significance was calculated using a two-sided Wilcoxon rank-sum test. **P*<0.05, ***P*<0.01, ****P*<0.001. **(g)** Pearson correlation heatmap of PDAC cell state abundances. Statistical significance was calculated using Pearson's correlation. *P<0.05, **P<0.01, ***P<0.001.

Given the strong link between cycling cell states and metastasis, we performed a multi-step differential expression analysis to identify candidate drivers of metastatic progression. We first identified 261 genes enriched in both cycling cell states (S3 and S6) relative to the other cell states, and also enriched in tumour cells compared with normal epithelial cells (Fig. 6.3a). Next, we filtered the 261 genes for those significantly associated with liver metastasis in both the Oxford-Kidani and the Link et al cohorts (Fig. 6.3a). This analysis yielded a 21-gene pre-metastatic (PreMet) signature, comprising genes enriched in primary tumours from patients who developed liver metastases, and maintained within cycling cell states at liver metastases (Fig. 6.3b). Among these genes, included high mobility group A1 (*HMGA1*) and glutathione peroxidase 4 (*GPX4*), both involved with fatty acid metabolism, a key feature of both cycling cell states (Fig. 6.2d). Additionally, the *KRT7* was significantly associated with metastasis, consistent with reports in PDAC[231]. Notably, when compared against each of the 21 genes individually, the combined PreMet score demonstrated superior predictive power (AUC = 0.83) for distant metastasis, emphasizing the importance of multiplex biomarker testing (Fig. 6.2b-c). Furthermore, the PreMet score was significantly enriched in both PT ($P = 3.0 \times 10^{-7}$) and MT ($P = 4.3 \times 10^{-9}$) tissue compared to NAT, while also significantly enriched in MT against PT tissue ($P = 4.4 \times 10^{-6}$) (Fig. 6.3d). Higher CNV burden strongly correlated with the PreMet score ($R = 0.43$, $P = 1.3 \times 10^{-6}$), suggesting that genomic alterations could be driving metastatic potential (Fig. 6.3e).

Lastly, the PreMet score was significantly associated with worse outcomes across PDAC datasets (Fig. 6.3f).

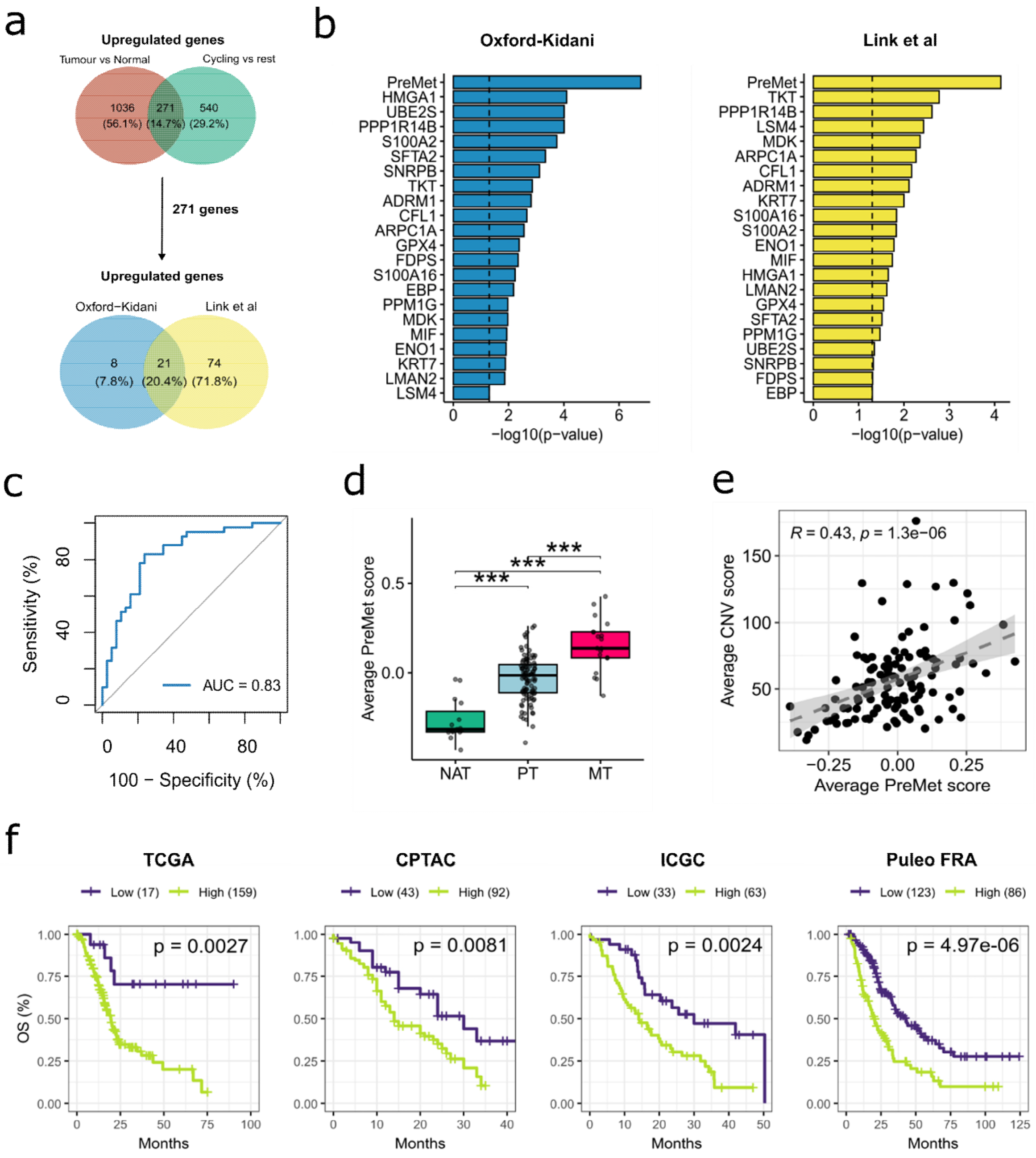


Figure 6.3 Identification of a metastasis-associated gene signature (PreMet). (a) Venn diagram of upregulated genes in tumour vs normal and cycling cell states (S3 and S6) against rest (top) yielding 271 genes. The 261 genes were further filtered for those significantly associated with distant metastasis in both the Oxford-Kidani and the Link et al cohorts (bottom). This yielded 21 metastasis-associated genes (PreMet). **(b)** Enrichment of PreMet gene levels and the average score in distant metastasis for both

Oxford-Kidani and Link et al cohorts. Statistical significance was calculated using a two-sided Wilcoxon rank-sum test. **(c)** ROC curve showing the predictive performance of the PreMet score in the Oxford-Kidani cohort. **(d)** Comparison of the PreMet score level across tissue types. Statistical significance was calculated using a two-sided Wilcoxon rank-sum test. *$P<0.05$, **$P<0.01$, ***$P<0.001$. **(e)** Pearson correlation plot of the PreMet score and the average CNV score. **(f)** Kaplan-Meier OS curves for the PreMet score across bulk PDAC datasets. Statistical significance was determined using the two-sided log-rank test.

6.4 Tumour cell heterogeneity and TME interactions

Tumour cell heterogeneity forms distinct interactions with the TME, driving disease progression, therapeutic resistance, and patient outcomes[224]. Therefore, we investigated the co-enrichment of the identified malignant cell states with CE abundances and key cell states. Notably, results show that each CE consisted of at least one malignant cell state, underscoring the connection between tumour heterogeneity and TME ecosystems (Fig. 6.4a).

The immunosuppressive/proliferative CE1 correlated with both proliferating cancer cell states (S3 and S6). This is consistent with cycling cells of several cell types forming "high-proliferation" niches[224]. Specifically, S3 malignant cells (G2/M) were correlated with the Mϕ_c2_CCL18 population, suggesting that this interaction may contribute to immunosuppression within this cellular niche. Therefore, to elucidate the underlying mechanisms, we investigated ligand-receptor interactions, focusing on both outgoing signals from cancer cell states, and incoming signals directed towards them (Fig. 6.4b). Interestingly, we identified the APOA1-TREM2 interaction originating from cycling cancer cells and targeting CE1 myeloid cells (CE1 Outgoing Interactions). *APOA1* encodes apolipoprotein A1, a key regulator of cellular cholesterol homeostasis that modulates cholesterol transport, while *TREM2* expression has been shown to be expressed in lipid-associated macrophages (LAMs), which is linked to the immunosuppressive M2 phenotype, and poor patient outcomes[232]. In addition, we observed elevated NECTIN2-TIGIT, and PVR-TIGIT checkpoint signalling from cycling cancer cells to CE1 CD4+ T and CD8+ T/NK cells, further underscoring immune evasive mechanisms within this CE, that are largely driven by metastatic cancer cells. On the other hand, we identified incoming SIRPA-CD47 signals originating from CE1 myeloid cells towards cycling cancer cells (CE1 Incoming Interactions). This “don’t eat me” signal enables cancer cells to evade phagocytosis[233]. Furthermore, we identified various signals from pericytes towards cycling cancer cells, including THY1-ADGRES, JAG1-NOTCH3, and SEMA5A-PLXNA1.

Furthermore, cells within the TLS-like CE2, including non-classical monocytes (Mon_c2_CD16) and cytotoxic CD8+ T cells (CD8T_c3_GZMH), correlated with the hypoxic cancer cell state (S5) (Fig. 6.4a). Outgoing signals from S5 cancer cells included HLA-A-CD8A/B and HLA-E-CD8A/B interactions with CD8+ T cells, suggesting enhanced antigen presentation and potential anti-tumour cytotoxic activity (CE2 Outgoing Interactions) (Fig. 6.4c). These findings are consistent with elevated MHC-I antigen presentation in tumour cells under hypoxic conditions, leading to antigen-specific cytotoxic T cell killing[234]. Additionally, we identified elevated BMP2-BMPR2 signalling towards endothelial cells, and this signalling axis has been reported to induce a proinflammatory endothelial phenotype[235]. Incoming signals included elevated IL33-IL1RAP from endothelial cells, and IL1B-IL1RAP from myeloid cells and stromal cells (CE2 Incoming Interactions), further highlighting the inflammatory phenotype of this CE.

Furthermore, CE3 correlated strongly, with the aggressive basal-like cancer phenotype (S1), as well as the PreMet signature score, suggesting that this CE represents the ‘pre-metastatic niche’. This CE included both incoming and outgoing signals involved in ECM degradation, including

MMP2-MMP14, and PLAU-PLAUR (Fig. 6.4d). Additionally, we observed elevated CXCL14-CXCR4 and MIF-CXCR4 signalling from basal-like cancer cells towards CE3 myeloid cells. We also identified increased SEMA3D-PLXNA1 originating from CE3 endothelial cells and directed towards basal-like cancer cells. This axis has previously been implicated in driving PDAC metastasis[236], thus confirming the pro-metastatic phenotype of the EC_c2_CCL21 population.

Furthermore, classical-like cancer cells correlated with CE4 TME cell states, including activated CD4+ T cells (CD4T_c1_ANXA1), and Mast_c2_LMNA cells (Fig. 6.4a). Most outgoing signals were directed towards CE4 endothelial cells, including JAG1-NOTCH3/4, and PDGFA-PDGFRB/A (CD4 Outgoing Interactions) (Fig. 6.4e). Notably, this CE displayed elevated AREG-ERBB2 signalling originating from CE4 B/Plasma, CD8+ T/NK, and myeloid cells towards classical-like cancer cells.

Lastly, the NAT-associated CE5 showed a strong correlation with progenitor cancer cells (S4), reflecting features of an early-stage tumour ecosystem (Fig. 6.4a). Progenitor cancer cells displayed enriched MHC-II-CD4 outgoing interactions towards DCs (Supplementary Fig. 6). Additionally, incoming communication from CE5 stromal cells included the FGF7-FGFR2 signalling axis.

Next, we calculated an average for each ligand-receptor signal and investigated their association with patient prognosis. Results show that the majority of signals associated with worse survival shared the ITGB1 receptor (Fig. 6.4c). ITGB1 overexpression has been shown to increase cancer cell adhesion to the extracellular matrix, and enhance invasiveness, thus facilitating tumour dissemination[237]. Conversely, elevated CXCL16-CXCR6 expression, which was specifically enriched in CE1, was associated with improved survival (Fig. 6.4c). This signalling axis has recently been reported to promote the survival of PD1^{+} TIM3^{+} terminally dysfunctional CD8$^+$ T cells, and CXCR6 deletion resulted in decreased tumour growth control[238]. Taken together, these findings highlight remarkable diversity of communication across cellular ecosystems, and suggest that malignant and TME cell states may coevolve during disease progression through coordinated bidirectional communications.

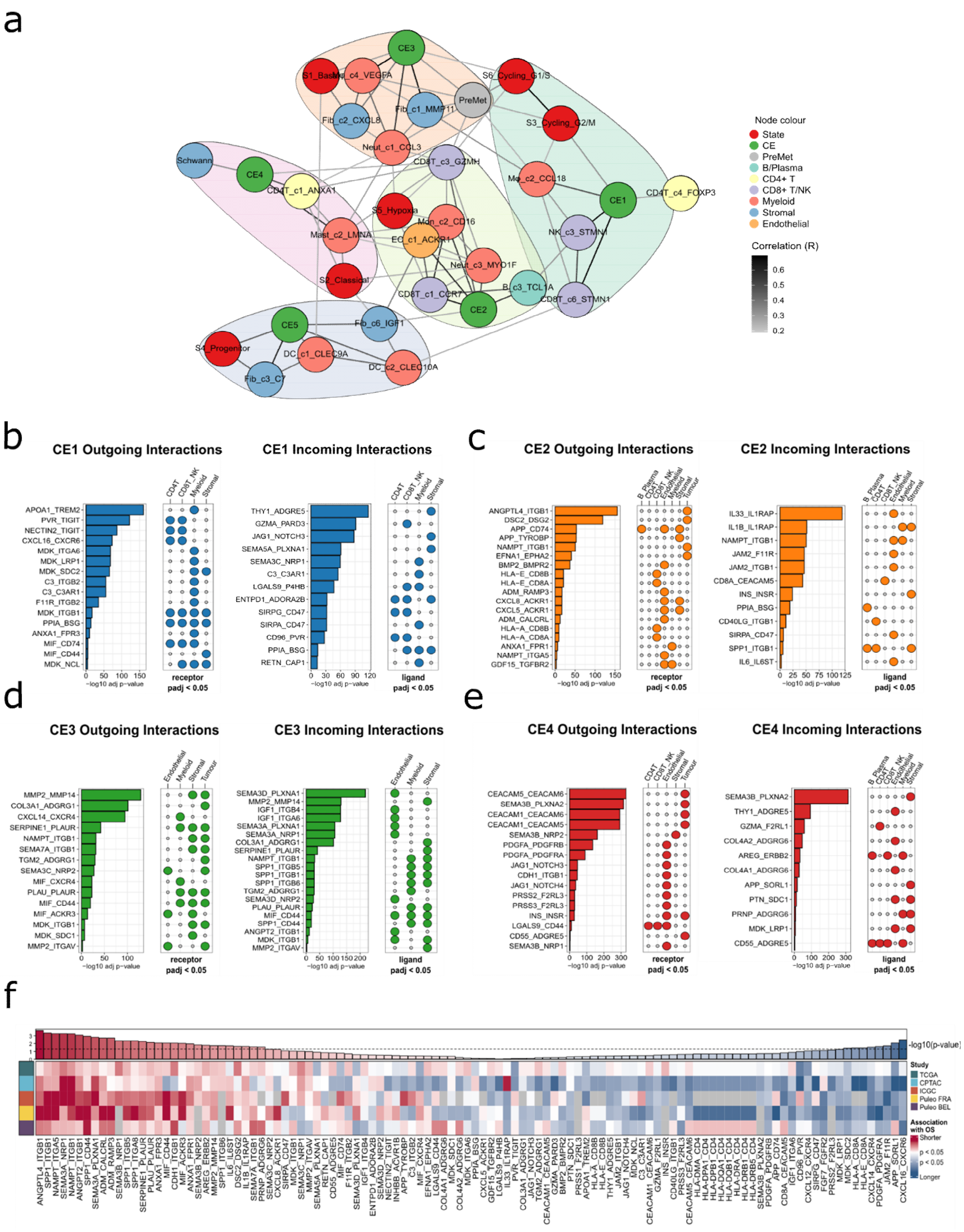


Figure 6.4 Malignant CE communication networks. (a) Network plot of five cellular communities, consisting of cancer cell states, CEs, and TME cell states. Nodes are coloured based on the feature or

cell type. Edge colour indicates the Pearson correlation coefficient. **(b-e)** Outgoing and incoming interactions between malignant cell states and TME cell types of the respective CE. The bar plot shows ligand-receptor pairs enriched in each CE, ranked by the -log10 (p-value), determined using a two-sided Wilcoxon rank-sum test. Dot plots show the enrichment of the receptor and ligand, for outgoing and incoming signals, respectively. Statistical significance was calculated using a two-sided Wilcoxon rank-sum test. **(f)** Heatmap showing the association of the joint ligand-receptor expression scores with OS through multivariate Cox regression modelling. Bar plot on top represents the average the survival association for each ligand-receptor pair.

6.5 Mutational characteristics of CEs

Next, we characterized the genomic landscape of PDAC to identify genetic alterations associated with each individual CET. Notably, tumour cellularity was significantly higher in CET3 compared to all other CETs (Fig. 6.5a). By adjusting for tumour purity, we investigated differences in broad genomic events across CETs. We observed a slightly higher TMB in CET1 and CET4 tumours, however TMB was low overall, consistent with the low frequency of TMB-high (TMB-H) tumours in PDAC[35] (Fig. 6.5a-b). Furthermore, the MSI score was significantly lower in CET3 compared to other CETs (Fig. 6.5a). Additionally, the MSI score positively correlated with CE2 and CE5 levels (Fig. 6.5b). Conversely, CET3, followed by CET4 and CET1 displayed the highest aneuploidy scores (Fig. 6.5a-b). Despite these associations, this dataset did not include MSI or HRR deficiency annotations, limiting our ability to directly evaluate how these genomic features relate to CE abundance.

Next, we investigated differences in somatic alterations of genes across CETs (Fig. 6.5c). Notably, CET3 displayed the highest mutation frequency of *KRAS* and RTK-RAS pathway alterations compared to other CETs (Fig. 6.5d-f). In contrast, CET2 and CET5 exhibited the highest levels of *KRAS* wild-type tumours (Fig. 6.5e). This is in line with the enriched frequencies of MSI-H tumours in *KRAS*^{WT} tumours compared to *KRAS*^{MUT} tumours[35]. Similarly, CET3 displayed the highest *TP53* mutation frequencies, while CET2 and CET5 displayed the highest *TP53*^{WT} tumours (Fig. 6.5e-f). Notably, CET4 was characterized by a significantly elevated frequency of *DNAH17* mutations, whereas CET1 showed increased mutations in *PDZRN4, NCAM1, APBA2, TENM2, KIF26B,* and *IRX2*, albeit at lower frequencies (Fig. 6.5d).

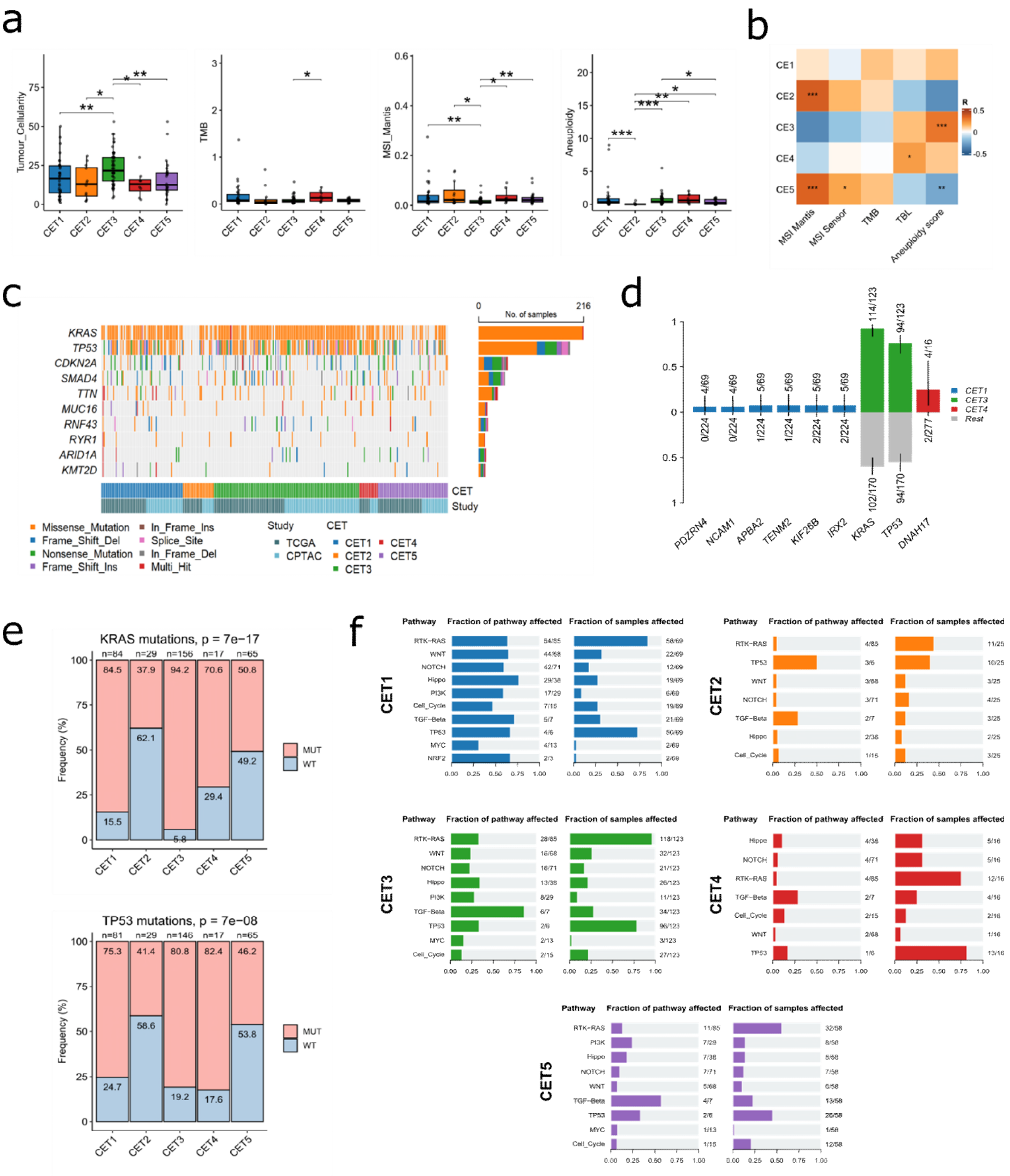


Figure 6.5 Mutational landscape characteristics of PDAC CETs. (a) Differences in tumour cellularity (%), TMB, MSI, and aneuploidy scores across CETs. The broad genomic alterations were adjusted by tumour purity. Statistical significance was determined using two-sided Wilcoxon rank-sum test. **(b)** Pearson correlation matrix of CE abundance and broad genomic alterations. **(c)** Oncoplot showing the

genetic alterations of the top 10 mutated genes across CETs in both TCGA and CPTAC cohorts. **(d)** Enrichment of mutational frequencies across CETs. Statistical significance was determined using the Fisher's exact test. **(e)** Frequencies of *KRAS* and *TP53* mutations across CETs. Statistical significance was determined using the two-sided Chi-squared test. **(f)** Mutational profiles of 10 canonical pathways for each CET. Frequencies of both mutation and affected samples in each canonical pathway are shown.

6.6 CNA characteristics of CEs

Moreover, we investigated broad and gene-specific CNAs across CEs to obtain functional insights of cancer cell state phenotypes. Consistent with its highest aneuploidy score, CET3 displayed the highest frequency of copy number amplifications and deletions across CETs (Fig. 6.6a). CET3 displayed significant deletions in chr18q, chr6q, chr12q, and chr17p, as well as amplifications in chr16, chr3q, chr12p, chr18q, and chr12p. Notably, CET5 displayed significantly elevated frequencies of chr5q amplifications, and this genomic alteration was associated with improved survival (Supplementary Fig. 7a). Furthermore, both CET2 and CET5 displayed elevated chr17p amplifications, while CET2 also displayed distinct deletion in chr2p, which contains genes involved in MMR including, *MSH2, MSH6*, and *EPCAM* (Fig. 6.6a).

Next, we explored gene-level copy number amplifications and deletions within the altered chromosomal regions. Focusing on oncogenes and tumour-suppressor genes curated from the OncoKB database[127], we evaluated their association with CETs at both CNA and gene expression levels. For instance, *KRAS* amplification on chromosome 12p was enriched in CET3, with *KRAS* expression positively correlating with CE3 frequencies (Fig. 6.6c). In contrast, expression was negatively correlated with CE2 and CE5, which were predominantly comprised of $KRAS^{WT}$ tumours (Fig. 6.5e). Furthermore, CET3 displayed amplifications of chr16p genes including, *MAPK3, PALB2,* and *AXIN1*. Additionally, the strongest amplification of genes was observed in chr3q, which included RTK-RAS signalling genes such as *PIK3CA/B* and *MAPK3K13,* as well *TP63*, which is a master regulator of the squamous PDAC subtype[239]. Additionally, amplification of chr3q was associated with worse survival (Supplementary Fig. 7b). In contrast, CET2 was enriched in amplification of chr17p genes, that consisted of multiple TSGs such as *MAP2K4, TP53,* and *GPS2*. These genes positively correlated with CE2 frequencies while they negatively correlated with CE3 frequencies, in line with the deletion of chr17p in CET3 (Fig. 6.5b-c).

Next, CET3 demonstrated significant deletions of chr18q genes, including *BCL2, SMAD4, GATA6,* and *SETBP1,* and these genes displayed a strong negative correlation with CE3 abundances (Fig. 6.5e). Given that GATA6 is an established marker of the classical phenotype, deletions may in part, drive the phenotypic switch to the basal-like cell state. Similarly, CET3 displayed deletions in chr6q genes, including *SESN1, REV3L*, and *ARID1B*. Overall, these findings reveal a strong association between genomic instability and PDAC CEs, suggesting their potentially utility as surrogate biomarkers in liquid biopsy approaches for patient stratification.

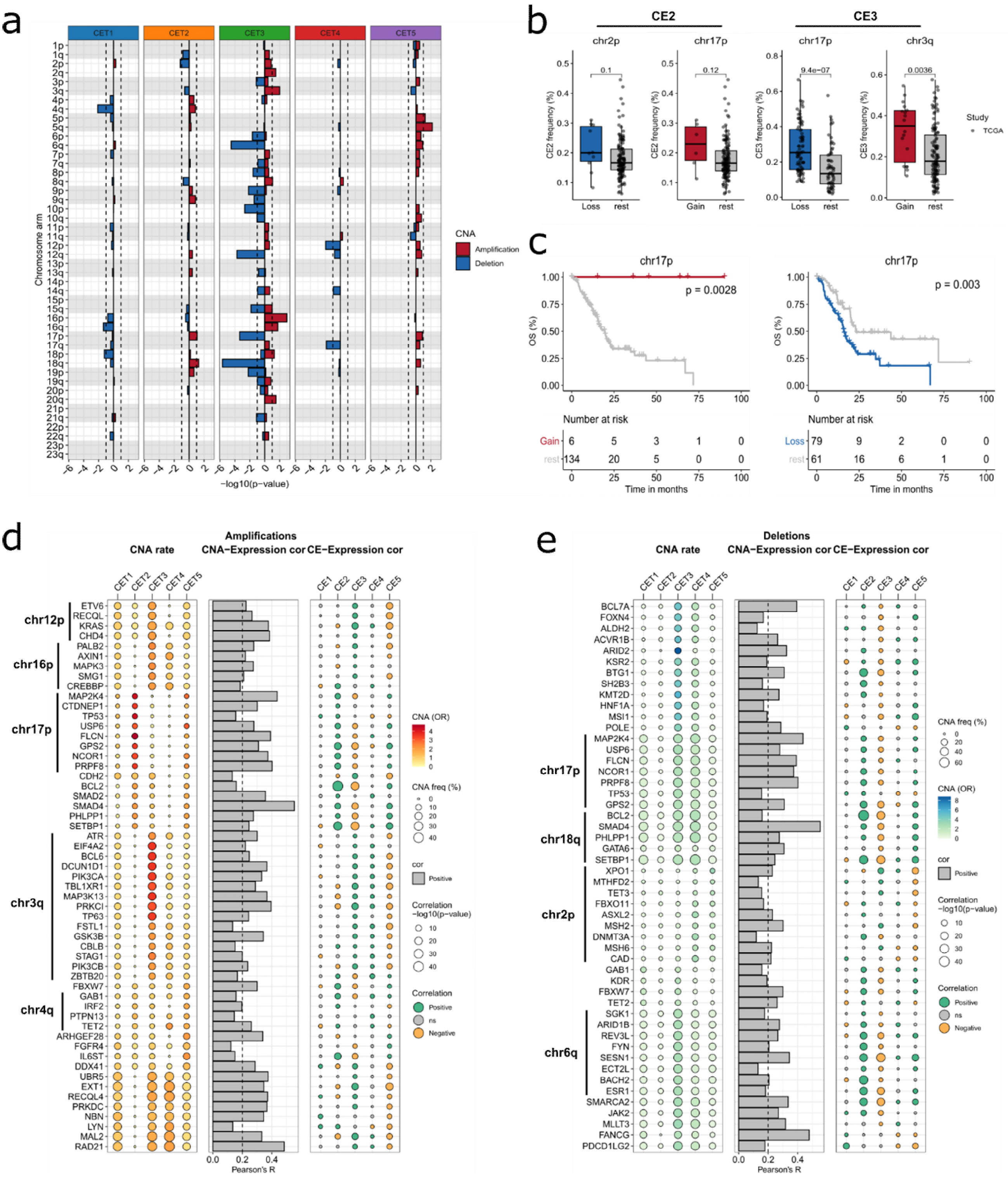


Figure 6.6 CNA characteristics of CEs. (a) Chromosomal arm copy number amplifications and deletions across CEs. Statistical significance was determined using the Fisher's exact test. **(b)** CE abundance and selected CNAs. Statistical significance was determined using a two-sided Wilcoxon rank-sum test. **(c)** Kaplan-Meier OS curves for the selected CNAs. Statistical significance was

determined using the two-sided log-rank test. **(d-e)** Gene-level copy number amplifications **(d)** and deletions **(e)** across CEs. Left dot plot shows the amplification **(d)** and deletion **(e)** rates of selected genes across CEs. Colour indicates the calculated odds ratio (OR) from Fisher's exact test. Size of the dot represents the amplification **(d)** and deletion **(e)** frequency in each CE. Middle bar plot shows the correlation between the CNA levels and expression for each gene. Right dot plot shows the correlation between the expression of the selected gene and CE abundance. Dots are coloured based on positive correlations (Pearson's R > 0.1 and p < 0.05), and negative correlations (Pearson's R > 0.1 and p < 0.05). Size of the dot indicates the -log10 (p-value). 6.6 CNA characteristics of malignant cell states.

6.7 CNA characteristics of tumour cell states

Since multiple cellular ecosystems, such as CE1 and CE3, have been shown to co-exist in individual tumours, this may mask the association of specific genomic alterations with distinct CEs. Therefore, we also compared CNAs across tumour cell states to identify those more precisely associated with disease progression. By investigating broad genomic alterations, we could see that cycling cell states (S3 and S6) belonging to CE1, displayed higher CNAs compared to the basal-like cell state belonging to CE3 (Fig. 6.7a). Results show that the basal-like cell state displayed the strongest amplification of chr3q, while it also displayed amplification in chr16p and chr16q, consistent with CET3 amplifications (Fig. 6.6a). Genes belonging to these chromosomal locations, including *PIK3CA, SOX2, BCL6,* and *TP63*, were significantly amplified in this malignant phenotype (Fig. 6.6b). Furthermore, both cycling cell states displayed enriched chr8q amplification, and deletions of chr17p. Specifically, the cycling cell states exhibited amplifications in chr8q genes, including *NDRG1*, *UBR5*, and *MYC*. Conversely, both the classical and progenitor cell states displayed enrichment of chr1q (Fig. 6.7a). This genomic alteration, was the most frequent (Supplementary Fig. 5c), suggesting that it represents an early event during disease progression that is maintained across cancer cells. Additionally, the classical cell state displayed elevated *GATA6* amplification, consistent with recent reports in classical PDAC subtype[42]. Furthermore, the cycling G2/M (S3) cell state displayed the strongest deletions of *CDKN2A* and *TP53*, suggesting that chr17p deletion is more frequent at later stages of tumour progression.

Next, we investigated gene amplifications and deletions shared between the PreMet score and cycling cell states to identify key drivers of PDAC metastasis. Results show that only amplifications in *MYC*, and *LPP* were shared across three phenotypes, whereas no common deletions were found (Fig. 6.7d). This suggests that metastatic dissemination in PDAC is largely driven by amplifications rather than deletions. Furthermore, *MYC* and *LPP* copy number alterations showed a strong association with patient survival, with copy number gains and amplifications showing worse survival (Fig. 6.7e).

In order to understand, how these cell states develop during disease progression, we performed distance-based phylogenetic tree analysis using the broad CNA alterations as input. Results show that the progenitor cell state with the lowest CNA burden (S5), was at the root of the phylogenetic tree, followed by the classical cancer cell state (S2), potentially by the acquisition of *GATA6* amplification (Fig. 6.7f). Next, the hypoxic cell state (S5) formed a separate branch, while the basal-like cell state evolved towards the cycling cell states, possibly through amplification of both *MYC,* and *LPP*, as well as deletions of TSGs, including *CDKN2A,* and *TP53*.

Given that MYC has been shown to regulate lipid metabolism by increasing APOA1 expression in prostate cancer[240], and that *APOA1*-mediated communication with *TREM2*$^{+}$ macrophages was specific to the cycling cancer cell states (Fig. 6.4b), we proposed a mechanistic link underlying the establishment of CE1 in metastatic tumours (Fig. 6.7g). Overall, these findings reveal that variation in malignant cell states is driven more by CNAs than by mutations, and that late-stage amplifications of *MYC* and *LPP* may promote cancer cell metastasis through enhanced proliferation.

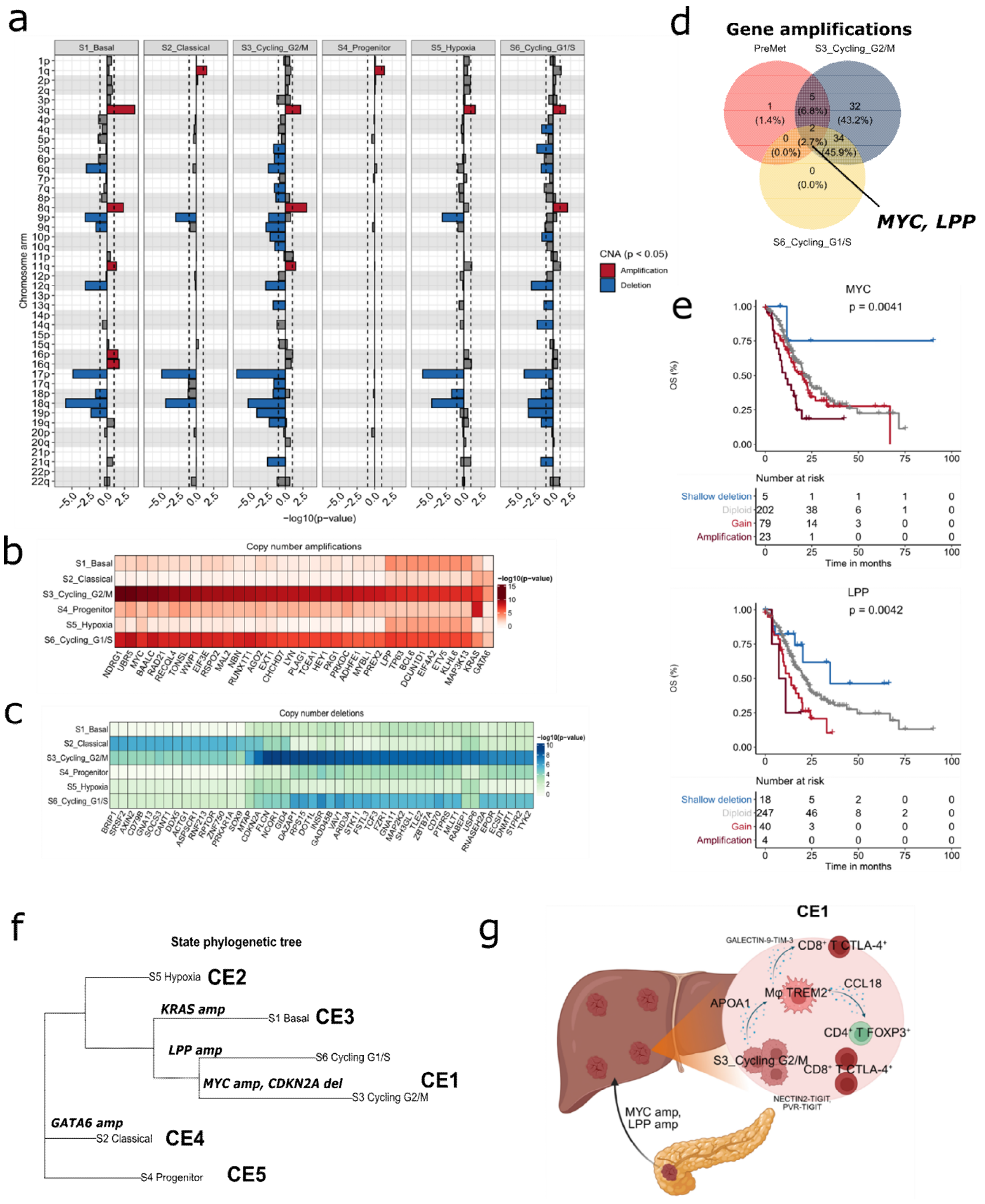


Figure 6.7 Genomic alterations of malignant cell states. (a) Chromosomal arm copy number amplifications and deletions across CEs. Statistical significance was determined using a two-sided Wilcoxon rank-sum test. **(b)** Heatmap of selected amplified genes across malignant cell states. **(c)** Heatmap of selected deleted genes across malignant cell states. **(d)** Venn diagram showing the shared amplified genes between the PreMet signature score and cycling cell states. **(e)** Kaplan-Meier OS curves for the selected CNAs. Statistical significance was determined using log-rank test. **(f)** Distance-based phylogenetic tree based on broad chromosomal alterations. **(g)** Schematic of malignant cell liver metastasis and establishment of CE1 through APOA1-TREM2 signalling axis.

6.8 Conclusions and Discussion

Tumour heterogeneity plays a central role in disease progression, treatment responses, as well as regulating the TME. However, the molecular mechanisms underlying the coordinated neoplastic and TME heterogeneity remain poorly understood. Our findings uncover distinct patterns of interactions between malignant cell states and the diverse cellular ecosystems of the TME, including specific immune and stromal cell populations.

Moreover, our results extend previous efforts to elucidate the complex crosstalk between tumour cells and the microenvironment. For instance, similar to the identified CE3, Oh et al. revealed the co-enrichment of the basal-like cancer cells with SPP1^{+} TAMs and myCAFs, reflecting the 'Moffitt Activated Stroma'[241]. However, due to their lower resolution analysis, key cell states driving tumour metastasis, such as hypoxic CAFs (Fib_c2_CXCL8), pro-tumourigenic neutrophils (Neut_c1_CCL3), and EC_c2_CCL21 were not characterized, thus limiting the identification of putative cell-cell interactions and potential therapeutic targets. Furthermore, our findings reveal the co-existence of proliferating malignant and immune cells forming the immunosuppressive CE1, a cellular ecosystem that remains poorly characterized in PDAC.

Various tumour-derived signals previously linked to cell proliferation, such as *PPIA* and *MDK*[242][243], may contribute to shaping the proliferative phenotype of CE1 cell states. Furthermore, given the strong association of CE1 with both chemotherapy resistance and ICB sensitivity, further investigation of this cellular ecosystem is of considerable clinical significance.

The relationship between genomic alterations, such as mutations and CNAs, and the formation of cellular ecosystems remains poorly understood in PDAC. Although PDAC has been shown to present limited mutational heterogeneity, with recurrent mutations in *KRAS, CDKN2A, TP53,* and *SMAD4*, our results show enrichment of $KRAS^{WT}$, and $TP53^{WT}$ tumours in immune-reactive CE2 and the normal stroma CE5. The enrichment of $KRAS^{WT}$ has also recently been reported in immune-reactive/fibrotic (IE/F) PDAC tumours[244]. A possible explanation is that these subtypes, displayed the lowest tumour cellularity, possibly masking mutational calls. PDAC is characterized by a pronounced fibrotic stroma and low tumour cellularity, limiting our ability to comprehensively profile the genomic landscape of individual tumours[245]. Therefore, purifying tumour cells using techniques such patient-derived organoids (PDOs) or laser capture microdissection (LCM) could offer a promising approach to investigate the mutational landscape of PDAC tumours across CETs.

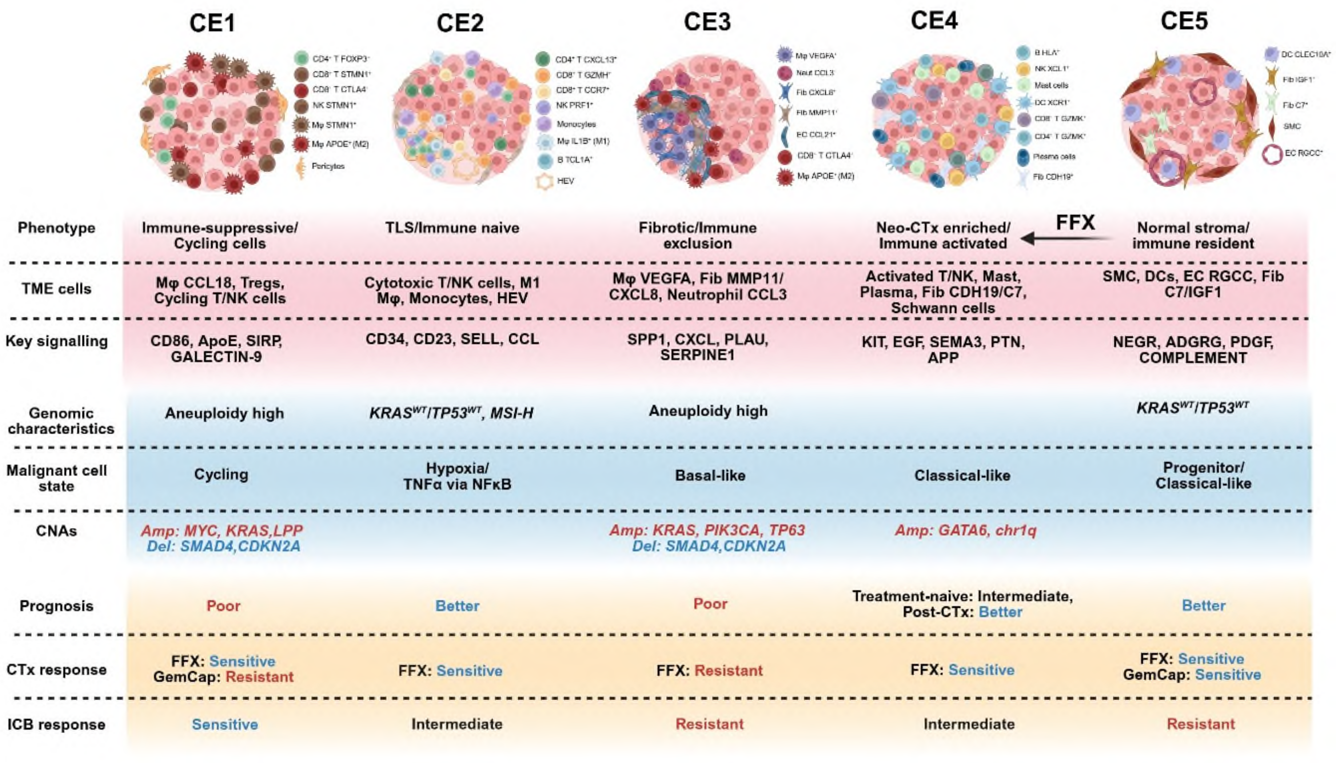
Furthermore, by adjusting for tumour cellularity, we revealed that CE2 was significantly associated with a higher MSI score. Given that MSI-H tumours are predominantly enriched in $KRAS^{WT}$ cases[35], this suggests that a subset of CE2-enriched patient tumours may represent true MSI-H cases. Additionally, we observed elevated TMB levels, normalized for tumour purity, in CET1 and CET4 tumours. Although these subtypes did not correlate with MSI, these findings are consistent with recently reported TMB-H/MSS PDAC tumours[246].

Genomic instability plays a significant role in tumour metastasis, and has been linked with transcriptomic subtypes in PDAC[43]. Consistent with previous reports, our results reveal mutant *KRAS* and *GATA6* amplifications in basal-like, and classical subtypes, respectively[42]. Furthermore, *KRAS* amplification was strongly associated with the fibrotic/angiogenic CET3, suggesting a mechanistic link between this genomic alteration and the basal-like phenotype, and consequently establishing this cellular ecosystem. An *in vitro* study of pancreatic cancer cells has shown that KRAS upregulates the secretion of PAI-1 (encoded by *SERPINE1*) via ERK signalling, which in turn activates pancreatic stellate cells (PSCs), leading to increased tumour stiffness and poor prognosis[247]. Given that basal-like cancer cells distinctly express this cytokine, together with the CAF_c2_CXCL8 population in CET3, it may play a central role in establishing this CE downstream of *KRAS* amplification.

Furthermore, our analysis showed that cycling cancer cells were more strongly associated with metastasis than the basal-like phenotype, and phylogenetic reconstruction indicated that they emerged after the basal-like state. This suggests that chromosomal alterations following mutant *KRAS* amplification are more strongly associated with metastatic spread. When compared with the PreMet score, results show that *MYC* and *LPP* amplifications are major determinants of metastasis in PDAC. Despite previous reports linking *MYC* amplification with metastatic burden in advanced PDAC, the association between *LPP* and PDAC metastasis remains largely unexplored[248]. LPP is a member of zyxin family of LIM proteins involved in mediating cellular responses to mechanical cues[249], and it has been reported to enable cancer cells to metastasise through the promotion of invadopodia formation. Therefore, further investigation of LPP amplification in PDAC is warranted, as it could provide key insights into tumour recurrence and metastasis.

To sustain their growth and survival, proliferating cancer cells rewire their metabolism, and our findings reveal enhanced lipid metabolism in metastatic cycling cells. Supporting this, a recent study in advanced prostate cancer demonstrated that *MYC* amplification upregulates APOA1 expression, linking cancer genetics to lipid metabolism and metastasis[240]. Similarly, we observed enriched APOA1-TREM2 communication in cycling cancer cell states with CE1 macrophages, aligning with the reported positive correlation between *APOA1* expression and M2 macrophages across multiple cancer types[250]. Interestingly, a recent study in glioblastoma demonstrated that targeting ApoA1 significantly enhanced CD8+ T cell proliferation, but only in the presence of TAMs[251]. This finding suggests that ApoA1-induced metabolic reprogramming in TAMs promotes anti-tumour immune responses mediated by CD8+ T cells. In contrast, in an ovarian cancer mouse model, cancer-derived ApoA1 was shown to bind ABC trasporters, facilitating cholesterol efflux and reducing lipid rafts in macrophages, ultimately driving TAM polarization and immunosuppression[252]. These discrepancies may reflect differences in the TME, thus warranting further investigation in PDAC metastasis.

Taken together, our findings reveal coordinated interactions between cancer cell states and cellular ecosystems, driven in part by underlying genomic alterations. Thus, beyond evaluating TME-derived cytokines as biomarkers for patient stratification, assessing genomic alterations, such as *MYC* or mutant *KRAS* amplifications could offer additional clinical value. Supporting this notion, a recent study identified an ecotype in sarcoma, analogous to CE1, that was resistant to chemotherapy yet exhibited superior responses to ICB and consisted of MYC amplified epithelial cells[222]. Collectively, applying this multilayered cancer and CE taxonomy could enable more accurate patient stratification for treatment selection, improve prediction of recurrence, and expand the pool of patients who may benefit from immunotherapy (Fig. 6.8).



	CE1	CE2	CE3	CE4	CE5
Phenotype	Immune-suppressive/ Cycling cells	TLS/Immune naive	Fibrotic/Immune exclusion	Neo-CTx enriched/ Immune activated	Normal stroma/ immune resident
TME cells	Mφ CCL18, Tregs, Cycling T/NK cells	Cytotoxic T/NK cells, M1 Mφ, Monocytes, HEV	Mφ VEGFA, Fib MMP11/ CXCL8, Neutrophil CCL3	Activated T/NK, Mast, Plasma, Fib CDH19/C7, Schwann cells	SMC, DCs, EC RGCC, Fib C7/IGF1
Key signalling	CD86, ApoE, SIRP, GALECTIN-9	CD34, CD23, SELL, CCL	SPP1, CXCL, PLAU, SERPINE1	KIT, EGF, SEMA3, PTN, APP	NEGR, ADGRG, PDGF, COMPLEMENT
Genomic characteristics	Aneuploidy high	$KRAS^{WT}$/$TP53^{WT}$, *MSI-H*	Aneuploidy high		$KRAS^{WT}$/$TP53^{WT}$
Malignant cell state	Cycling	Hypoxia/ TNFα via NFκB	Basal-like	Classical-like	Progenitor/ Classical-like
CNAs	*Amp: MYC, KRAS,LPP* *Del: SMAD4,CDKN2A*		*Amp: KRAS, PIK3CA, TP63* *Del: SMAD4,CDKN2A*	*Amp: GATA6, chr1q*	
Prognosis	Poor	Better	Poor	Treatment-naive: Intermediate, Post-CTx: Better	Better
CTx response	FFX: Sensitive GemCap: Resistant	FFX: Sensitive	FFX: Resistant	FFX: Sensitive	FFX: Sensitive GemCap: Sensitive
ICB response	Sensitive	Intermediate	Resistant	Intermediate	Resistant

Figure 6.8 Multilayered characterization of PDAC CEs. Schematic representation of PDAC multi-cellular CEs. Top bar outlines TME characteristics, while middle bar outlines genomic alterations and malignant cell states specific to each CE. Final bar outlines the prognosis and treatment responses for each CE.

Chapter 7

Summary and future work

7.1 Overall summary

By integrating publicly available single-cell and bulk RNA-seq, spatial transcriptomics, and in-house proteomics datasets, this multi-omics approach enabled a comprehensive characterization of the PDAC cellular architecture across three key aspects:

- The establishment of a high-resolution single-cell atlas of PDAC, providing a detailed taxonomy of tumour and TME cell states.
- The identification of distinct multi-cellular ecosystems characterized by specific spatial organizations, intercellular communication networks, patient prognoses, and therapeutic responses.
- The delineation of relationships between tumour cell heterogeneity, underlying genomic alterations, and the potential establishment of the identified CEs.

7.2 Single-cell landscape of PDAC

While previous studies have described the cellular landscape within PDAC, the limited number of samples and variations in experimental design limit their capacity to serve as a universal reference for this disease. Additionally, the quality of integrated data depends on the choice of data integration method, with tools such as Seurat's RPCA[98], and Harmony[99] failing to separate clusters in atlas-level studies[253]. In this thesis, using the top-performing integration tool scANVI, we integrated all major scRNA-seq PDAC studies published until recently, consisting of various tissue types, and treatment conditions. The 53 TME, along with 5 non-malignant epithelial, and 6 tumour cell states, could represent consensus annotation of the cellular landscape of PDAC. Importantly, the large-scale atlas enabled the identification of previously uncharacterized cell types such as the *CDH19*$^{+}$ fibroblast population (Fib_c4_CDH19), that represented a subset of both the iCAF and neurotropic cell states[56 61]. Nevertheless, certain immune subsets such as γδ T cells and CD8+ Tregs were not identified as separate clusters, highlighting the limitation of the single-cell atlas for characterizing rare immune cell subsets. Additionally, only Zhang et al. included matched primary tumours and liver metastases, limiting our ability to perform paired tracking of the TME cell states across tissue types. However, mapping additional datasets to the atlas using tools such as scArches[168], could facilitate the identification of additional subsets and improve cross-tissue comparisons.

By harmonizing clinical information, we were able to perform association analysis of cell state abundances with covariates such as tissue types, neo-CTx treatment, age, gender, and histological grade. Consistent with previous reports[143], our findings indicate that the metastatic tissue-enriched Mϕ_c2_CCL18 population plays a central role in mediating immunosuppression in PDAC, characterized by the high expression of Treg recruitment chemokines (*CCL18* and *CCL20)*, and multiple checkpoint molecules, including *LGALS9* (encoding Galectin-9), *SIRPA,* and *LAIR1*. Interestingly, this population exhibited similar signatures, including 'phagocytosis', and the 'M2 signature' with the Mϕ_c3_SLC40A1. However, only the Mϕ_c2_CCL18 cell state was enriched in metastatic tissue and showed distinct upregulation of the 'lipid metabolism' and 'checkpoint ligands' signatures. These results highlight the complexity of macrophage heterogeneity, and suggest that metabolic reprogramming within distinct TMEs, such as enhanced lipid metabolism, may promote immunosuppressive mechanisms that facilitate cancer progression. Additionally, these populations displayed opposing responses to neo-CTx, with the metabolically active Mϕ_c2_CCL18 decreasing after neo-CTx, whereas the Mϕ_c3_SLC40A1 was enriched. However, the mechanisms underlying these differences remain poorly understood and require further investigation. The depletion of Mϕ_c2_CCL18 population following neo-CTx was consistent with observations in responding CRC patients, whereas those with progressive disease exhibited enrichment of this population after treatment[254]. One possible explanation is that residual cancer cells following neo-CTx may metabolically reprogram the enriched Mϕ_c3_SLC40A1 population into Mϕ_c2_CCL18 cells, thereby enhancing immunosuppression and promoting metastasis.

Furthermore, the single-cell atlas enabled the assessment of cell-cell communication dynamics across tissue types (PT vs MT), and treatment conditions (treatment-naïve vs neo-CTx treated). By treating individual studies as independent variables, we identified robust communication networks consistently observed across datasets, providing reliable signals that can be further validated experimentally. Taken together, this single-cell atlas may serve as a valuable resource for the pancreatic cancer scientific community and may help accelerate future discoveries. Ongoing efforts aim to incorporate additional datasets to capture a broader spectrum of cell states and further enhance the quality of the atlas as a comprehensive reference. This atlas is publicly available as a resource for the scientific community (https://trainingidn.shinyapps.io/demetriou_pdac_atlas_shinyApp/).

7.3 PDAC CEs and their clinical significance

In this thesis, we leveraged the single-cell atlas comprising a large number of patient samples and applied a multimodal approach to identify robust cellular communities present in PDAC across diverse clinical conditions. Importantly, this analysis was based on absolute cell frequencies, thereby overcoming the limitations associated with deconvolution methods applied in bulk datasets, such as the inability to resolve transcriptionally similar cell states[255]. Furthermore, we validated the colocalization and aggregation of cell states within CEs using spatial transcriptomics, as well as the overlap with independent morphological-based clustering using stLearn[110]. Although no statistically significant concordance was derived between the CE and Morph clusters, further analysis incorporating additional ST datasets, such as the recently published study including normal pancreas, primary tumours, and liver metastases[256], would be valuable to strengthen these observations.

Additionally, the co-existence of CEs in individual tumours aligns with the “reactive” and “deserted” subTMEs previously reported in PDAC[64]. Specifically, the deserted subTMEs shared features with the normal stroma CE5, lacking activated fibroblasts and tumour suppressive features, whereas the reactive subTME resembled multiple CEs, including the fibrotic CE3 and the immune-enriched CE1 and CE2. Notably, the identified CEs were distinct from previously reported TME-based subtypes, therefore, our approach captures a broader spectrum of cellular communities and provides new insights into TME heterogeneity of PDAC.

Furthermore, we developed a new deconvolution tool capable of accurately investigating the associations of cell state and CE abundances with patient prognosis and treatment response from bulk transcriptomic data. Both CTcaller and CScaller tools outperformed existing deconvolution methods, including CIBERSORTx, InstaPrism, MuSiC, and BisqueRNA. However, additional benchmarking across independent datasets will be necessary to further validate the accuracy of our approach. Both individual cell states and CEs associated with patient outcomes across independent datasets, validating the importance of the TME in tumour progression and response to therapies. Specifically, we identified a strong inverse association between the complement-enriched Fib_c3_C7 and the hypoxic Fib_c2_CXCL8 populations with patient survival, suggesting that these CAF states may serve as prognostic biomarkers. Additionally, the myCAF-like Fib_c1_MMP11 population was also associated with worse outcomes. However, a previous mouse model study depleting α-SMA+ CAFs resulted in enhanced PDAC progression, reduced survival and immunosuppression[257], suggesting that myofibroblasts are tumour-restraining. Our findings show that α-SMA (encoded by *ACTA2*) was highly expressed in SMCs, belonging to CE5 and associated with improved outcomes. Thus, more accurate markers are needed to distinguish CAF populations to enable patient stratification.

Furthermore, our findings reveal variable associations of CEs with systemic chemotherapy dynamics and responses. For instance, CE1 which includes multiple proliferating cell types was significantly depleted following FFX neo-CTx in both the single-cell atlas and a bulk validation dataset, indicating sensitivity to this treatment regime. In contrast, CE1 displayed opposite response to GemCap adjuvant chemotherapy, with this resistance largely driven by the proliferating macrophages (Mϕ_c5_STMN1), potentially through increased deoxycytidine (dC) production, thus reducing gemcitabine uptake. These findings underscore the clinical utility of CE1 as a dynamic biomarker for predicting chemotherapy response. For instance, patients who relapse following FOLFIRINOX, due to the re-establishment of a pre-existing CE1 or a *de novo* CE1 development, are unlikely to benefit from second-line gemcitabine-based therapy. To further support this, multiple studies have reported improved survival outcomes with neo-adjuvant FFX compared to neo-adjuvant gemcitabine-based chemotherapy in metastatic PDAC, but not in resectable or borderline resectable PDAC[258] [259]. Given that CE1 is more enriched in metastatic tumours compared to resectable PDAC, where it accounts for only 18.9%, this could partly explain the superior efficacy of FFX over gemcitabine in the metastatic setting. Conversely, patients with CE5-enriched tumours exhibited greater sensitivity to gemcitabine-based chemotherapy, further emphasizing the importance of stratifying patients based on their CE profiles to maximize therapeutic efficacy and reduce treatment-related toxicity.

PDAC patients are largely refractory to ICB treatment, and there is an urgent need to identify more accurate biomarkers for selecting patients that would benefit to immunotherapy. The presence of TLSs has been shown to act as a predictive biomarker for immunotherapy effectiveness across cancer types[260]. For instance, a retrospective study in PDAC has shown that mature TLSs predicted favourable response to anti-PD-L1, independent of PD-L1 expression status and CD8$^+$ T cell density[261]. However, contradictory findings have reported that low TLS scores were associated with better responses to both monotherapy and combination ICB regimens[262]. This discrepancy could be explained by the observed functional heterogeneity of TLSs, as immature TLSs (imTLSs) have been associated with reduced presence of DCs, and enrichment of exhausted T cells, underlying an immunosuppressive TME[263]. Conversely, mature TLSs contain Tfh cells, and high endothelial venules (HEVs), features associated with enhanced anti-tumour immunity and improved ICB responses. In this thesis, CE1 abundance outperformed other biomarkers, including TLS levels, for predicting ICB response across multiple cohorts. These findings were particularly intriguing, as CE1 did not exhibit the canonical immune-hot phenotype. Instead, CE1 was characterized by immunosuppressive cell states, including the Mϕ_c2_CCL18 population, Tregs (CD4T_c4_FOXP3), exhausted CD8+ T cells (CD8T_c2_CTLA4), as well as proliferating macrophages and T/NK cells. Similarly, a pan-cancer study revealed an IFNγ-associated cellular ecosystem, composed of APOE$^+$ macrophages and Tregs, which exhibited superior response to ICB[264]. However, other studies have reported that TREM2$^+$ TAMs are associated with worse immunotherapy outcomes across cancer types[265] [266]. This discrepancy could be explained by the phenotypic similarity of Mϕ_c2_CCL18 with both

Mϕ_c3_SLC40A1 and Mϕ_c4_VEGFA subsets, the latter of which we found to be strongly associated with reduced ICB efficacy.

Furthermore, our findings suggest that the proliferating CD8$^+$ T and NK cells of this CE were the key drivers for ICB response. The cycling CD8T_c6_STMN1 cell state expressed both activation (*GZMH*) and exhaustion (*LAG3*) markers, consistent with the enriched dysfunctional signature and tumour reactivity of proliferating CD8$^+$ T cells[267]. Additionally, proliferating exhausted CD8$^+$ T cells (P-Tex) were recently reported in antigen-presenting cell (APC) niches in HNSCC, and provided persistent anti-tumour effects[268]. Although DCs were not assigned to CE1, we found a positive correlation of the DC_c2_CLEC10A population with the cycling CD8T_c6_STMN1 cell state, indicating enhanced antigen presentation in this CE. Further investigation of the molecular mechanisms underlying this proliferating phenotype, as well as its spatial colocalization with APCs such as DCs, is warranted.

In line with recent findings, we observed enrichment of effector T cells and a reduction in immunosuppressive signalling following neo-adjuvant FFX chemotherapy, which was associated with prolonged patient survival in PDAC. However, FFX treatment also significantly depleted the cycling CD8$^+$ T and NK cell populations, suggesting that FFX may dampen rather than enhance the efficacy of ICB when given in combination or as subsequent treatment. Taken together, these findings confirm the clinical relevance of the identified CEs, and indicate CE1 as a key biomarker associated with chemotherapy resistance yet enhanced sensitivity to ICB.

7.4 Tumour cell heterogeneity and PDAC CEs

Our findings reveal distinct communications between tumour cell states and TME cells, further validating that these compartments coevolve during PDAC progression. These results are consistent with previous reports linking the aggressive basal-like phenotype with *SPP1*$^+$

macrophages and myCAFs[241 269]. However, our analysis further refines the landscape of cancer-TME interactions, identifying key communication signals, such as the pro-fibrotic *SERPINE1* in basal-like cancer cells (S1) contributing in part to the establishment CE3, whereas the TNFα via NFκB/hypoxic cell state (S5) exhibited genes associated with antigen presentation (including *HLA-A* and *HLA-E)* in the inflammatory CE2. Moreover, the metastatic cycling cells (S3 and S6) expressed high levels of immune checkpoint molecules, including *NECTIN2, PVR,* and *CD47,* thus promoting the immunosuppressive CE1.

High CNA burden has been linked to aggressive phenotypes and worse survival in resectable and metastatic PDAC[270]. Our results link somatic CNAs of oncogenes and TSGs with tumour cell heterogeneity, leading in part, to the subsequent establishment of certain CEs. For instance, we identified unique chromosomal arm alterations in CET2 tumours, including amplification of 17p, which was associated with improved patient survival. Notably, although *TP53*, located in chromosome 17p, showed higher amplification rates in CET2, this subtype exhibited higher proportions $TP53^{WT}$ cases. This suggests enhanced *TP53* activity within CET2 tumours and points to a potential immunomodulatory role. This is consistent with the immunogenic consensus subtype (CMS1) of CRC, which exhibits the highest $TP53^{WT}$ cases and comprises the majority of MSI tumours[271]. Interestingly, this study revealed that within CMS1 tumours, $TP53^{MUT}$ exhibited significantly fewer cytotoxic $CD8^+$ T cells than $TP53^{WT}$, but this was not seen other CMS groups. Furthermore, *TP53* expression has been shown to be upregulated following interferon signalling, which can be triggered by both oncogenic stress and viral infection[272 273]. Therefore, these findings suggest that *TP53* mutation status may play a critical role in maintaining effective anti-tumour immunity particularly in CE2 tumours, and could serve as a potential biomarker for identifying CE2 tumours in PDAC. Further analysis comparing $TP53^{MUT}$ and $TP53^{WT}$ subgroups within CE2 may reveal clinically meaningful findings.

Furthermore, our analyses link *MYC* amplifications, distinctly enriched in proliferating tumour cells, with the immunosuppressive/proliferating CE1. These findings suggest that these MYC-activated tumour cells may induce the proliferation of other cell types in their local TME through specific intercellular interactions. One possible key interaction is APOA1-TREM2, as APOA1 expression has been reported to be upregulated by MYC activation[240], while TREM2 signalling in macrophages is known to enhance their proliferation and survival[274]. Additionally, MYC has been shown to directly upregulate expression of the checkpoint molecules *CD47*, a critical "don't eat me" ligand, as well as *PD-L1*[275]. Beyond mediating phagocytosis evasion through the CD47-SIRPA axis, our results reveal an additional interaction between proliferating tumour cells and CE1 lymphocytes (CD4+ T, CD8+T, and NK cells) through CD47-SIRPG, and this interaction has been shown to drive antigen-specific T cell proliferation[276 277]. Therefore, these observations suggest that targeting SIRPA, rather than CD47, could inhibit tumour cell protection from phagocytosis while preserving CD8+ T cell proliferation mediated by CD47-SIRPG. Moreover, combining SIRPA blockage with therapies such as anti-TREM2 or anti-PD-L1 could further enhance immunotherapy efficacy in CE1-enriched tumours. Future mechanistic experiments in PDAC preclinical models are warranted to validate these intercellular interactions and inform the design of prospective combination therapies.

Chapter 8

Appendix

Study	Source	Sequencing Technology	Individuals	Tissue type	Neo-CTx
Steele et al[68]	GSE155698	10x v3	19	16 PTs, 3 NATs	N/A
Zhang et al[86]	GSE197177	10x v2	4	3 PTs, 1 NATs, 4 MTs	N/A
Werba et al[87]	GSE205013	10x v3	27	17 PTs, 10 MTs	1 CRT, 4 FFX, 2 Comb, 1 GP
Zhou et al[88]	phs002371.v1.p1	10x v3	17	17 PTs	1 CRT, 5 FFX, 1 Comb, 3 GP
Lin et al[89]	GSE154778	10x v2	16	10 PTs, 6 MTs	N/A
Moncada et al[90]	GSE111672	InDrop	2	2 PTs	N/A
Oh et al[91]	GSE231535	10x v3	2	2 PTs	N/A
Peng et al[92]	GSA: CRA001160	10x v2	35	24 PTs, 11 NATs	N/A
Xue et al[93]	GSE202742	BD Rhapsody	4	4 PTs	N/A

Supplementary Table 1 Single-cell atlas study summary. Study summary showing the number of individuals and samples from different tissue types and neo-CTx. Tissue type numbers are higher than individuals as certain individuals had more than one sample. CRT: Chemo-Radiotherapy; FFX: FOLFIRINOX; GP: Gemcitabine Nab-Paclitaxel; Comb: FOLFIRINOX and Gemcitabine Nab-Paclitaxel

cluster	gene	avg_log2FC	p_val_adj	pct.1	pct.2
CD4T_c1_ANXA1	KLRB1	1.21	3.17E-171	0.614	0.35
CD4T_c1_ANXA1	ANXA1	1.18	5.01E-204	0.786	0.512
CD4T_c1_ANXA1	IL7R	1.16	0	0.924	0.765
CD4T_c1_ANXA1	RORA	0.64	4.25E-46	0.558	0.446
CD4T_c1_ANXA1	MYADM	0.58	5.44E-34	0.541	0.432
CD4T_c2_CCR7	CCR7	1.95	0	0.618	0.191
CD4T_c2_CCR7	SELL	1.79	8.19E-255	0.517	0.173
CD4T_c2_CCR7	TCF7	1.41	8.22E-253	0.602	0.262
CD4T_c2_CCR7	KLF2	1.19	0	0.814	0.47
CD4T_c2_CCR7	AREG	0.93	8.70E-66	0.386	0.226
CD4T_c3_CXCL13	CXCL13	5.25	0	0.516	0.028
CD4T_c3_CXCL13	TOX2	3.26	7.75E-248	0.383	0.051
CD4T_c3_CXCL13	TOX	2.14	6.10E-201	0.464	0.107
CD4T_c3_CXCL13	PDCD1	1.76	3.06E-118	0.381	0.112
CD4T_c3_CXCL13	TIGIT	1.57	9.54E-202	0.711	0.264
CD4T_c4_FOXP3	FOXP3	5.26	0	0.502	0.021
CD4T_c4_FOXP3	IL2RA	4.44	0	0.508	0.054
CD4T_c4_FOXP3	TNFRSF4	3.97	0	0.604	0.124
CD4T_c4_FOXP3	TNFRSF9	3.57	0	0.391	0.037
CD4T_c4_FOXP3	CTLA4	3.10	0	0.754	0.168
CD4T_c5_GZMK	GZMK	3.46	0	0.88	0.174
CD4T_c5_GZMK	CCL4	3.12	0	0.739	0.171
CD4T_c5_GZMK	CST7	2.16	0	0.787	0.305
CD4T_c5_GZMK	CCL5	2.01	0	0.93	0.387
CD4T_c5_GZMK	NKG7	1.92	0	0.619	0.137
CD4T_c6_ISG15	GZMB	4.40	0	0.581	0.062
CD4T_c6_ISG15	CXCL13	3.35	4.92E-160	0.299	0.036
CD4T_c6_ISG15	IFI44L	3.20	9.69E-183	0.425	0.08
CD4T_c6_ISG15	IFI6	2.63	1.77E-172	0.6	0.2
CD4T_c6_ISG15	ISG15	2.36	3.55E-134	0.698	0.325
CD4T_c7_NKG7	NKG7	3.92	0	0.833	0.163
CD4T_c7_NKG7	GZMH	3.60	0	0.461	0.072
CD4T_c7_NKG7	GZMB	3.44	0	0.422	0.052
CD4T_c7_NKG7	CCL5	2.30	0	0.965	0.428
CD4T_c7_NKG7	GZMA	2.15	0	0.829	0.321

Supplementary Table 2 Top 5 differentially expressed genes across CD4+ T cell states.

cluster	gene	avg_log2FC	p_val_adj	pct.1	pct.2
CD8T_c1_CCR7	CCR7	3.24	0	0.578	0.08
CD8T_c1_CCR7	SELL	2.91	0	0.643	0.094
CD8T_c1_CCR7	TCF7	1.98	1.38E-216	0.7	0.218
CD8T_c1_CCR7	KLF2	1.33	1.92E-160	0.871	0.443
CD8T_c1_CCR7	IL7R	1.16	2.11E-175	0.97	0.463
CD8T_c2_CTLA4	CXCL13	4.58	0	0.378	0.024
CD8T_c2_CTLA4	CTLA4	4.13	0	0.641	0.055
CD8T_c2_CTLA4	IL2RA	3.97	1.26E-160	0.218	0.02
CD8T_c2_CTLA4	TNFRSF4	3.16	2.44E-100	0.203	0.041
CD8T_c2_CTLA4	TNFRSF9	2.54	2.07E-182	0.336	0.057
CD8T_c3_GZMH	GZMH	1.46	3.55E-240	0.649	0.379
CD8T_c3_GZMH	GZMK	1.38	0	0.774	0.414
CD8T_c3_GZMH	CST7	0.72	1.29E-205	0.895	0.732
CD8T_c3_GZMH	CCL5	0.47	4.62E-141	0.994	0.905
CD8T_c3_GZMH	TIGIT	0.30	1	0.313	0.302
CD8T_c4_GZMK	GZMK	1.82	0	0.955	0.393
CD8T_c4_GZMK	MYADM	0.91	1.54E-152	0.696	0.467
CD8T_c4_GZMK	TCF7	0.71	4.57E-44	0.332	0.215
CD8T_c4_GZMK	KLF2	0.63	1.63E-108	0.658	0.42
CD8T_c4_GZMK	IL7R	0.51	5.06E-111	0.695	0.439
CD8T_c5_KLRC1	IL7R	1.76	0	0.834	0.372
CD8T_c5_KLRC1	ANXA1	0.72	4.77E-133	0.875	0.663
CD8T_c5_KLRC1	RORA	0.52	1.23E-52	0.512	0.356
CD8T_c5_KLRC1	MYADM	0.51	6.34E-35	0.592	0.474
CD8T_c6_STMN1	CXCL13	3.17	5.40E-165	0.313	0.038
CD8T_c6_STMN1	CTLA4	1.75	9.69E-89	0.341	0.084
CD8T_c6_STMN1	IFI6	1.71	4.49E-101	0.618	0.231
CD8T_c6_STMN1	PDCD1	1.11	1.48E-49	0.305	0.096
CD8T_c6_STMN1	IFI44L	1.09	6.70E-27	0.218	0.08

Supplementary Table 3 Top 5 differentially expressed genes across CD8+ T cell states.

cluster	gene	avg_log2FC	p_val_adj	pct.1	pct.2
NK_c1_ISG15	GZMA	2.15	7.13E-97	0.893	0.789
NK_c1_ISG15	IFI44L	2.03	5.07E-75	0.273	0.077
NK_c1_ISG15	IFI6	2.02	4.31E-130	0.605	0.228
NK_c1_ISG15	ISG15	1.70	5.98E-84	0.7	0.361
NK_c1_ISG15	AREG	1.27	2.16E-77	0.66	0.378
NK_c2_PRF1	GZMB	1.96	0	0.799	0.398
NK_c2_PRF1	NKG7	1.80	0	1	0.851
NK_c2_PRF1	CST7	1.25	0	0.875	0.742
NK_c2_PRF1	KLRB1	1.23	0	0.786	0.411
NK_c2_PRF1	KLF2	0.97	2.54E-151	0.655	0.423
NK_c3_STMN1	GZMA	0.99	1.01E-32	0.859	0.791
NK_c3_STMN1	IFI6	0.74	6.94E-26	0.466	0.235
NK_c3_STMN1	TOX	0.73	2.38E-25	0.357	0.157
NK_c3_STMN1	AREG	0.56	3.49E-05	0.501	0.384
NK_c3_STMN1	ISG15	0.32	1.40E-05	0.534	0.369
NK_c4_XCL1	AREG	2.49	0	0.734	0.338
NK_c4_XCL1	KLRB1	1.47	0	0.845	0.411
NK_c4_XCL1	CCL4	0.74	4.01E-37	0.819	0.772
NK_c4_XCL1	NKG7	0.48	1.03E-167	1	0.855
NK_c4_XCL1	MYADM	0.32	3.37E-16	0.55	0.495

Supplementary Table 4 Top 5 differentially expressed genes across NK cell states.

cluster	gene	avg_log2FC	p_val_adj	pct.1	pct.2
B_c1_TNF	TNF	1.48	1.26E-51	0.274	0.152
B_c1_TNF	ID3	1.39	8.11E-95	0.512	0.341
B_c1_TNF	TUBA1A	0.93	1.83E-103	0.573	0.398
B_c1_TNF	DDIT4	0.74	3.51E-31	0.53	0.462
B_c1_TNF	INTS6	0.70	1.96E-42	0.554	0.466
B_c1_TNF	FCER2	0.54	1.22E-11	0.246	0.186
B_c2_NR4A3	NR4A2	1.19	7.11E-269	0.8	0.522
B_c2_NR4A3	SLC2A3	1.10	1.23E-134	0.594	0.378
B_c2_NR4A3	CREM	1.06	2.85E-158	0.648	0.421
B_c2_NR4A3	ZNF331	1.01	2.13E-131	0.638	0.423
B_c2_NR4A3	CD69	0.93	3.00E-121	0.688	0.492
B_c3_TCL1A	TCL1A	1.76	3.92E-121	0.422	0.193
B_c3_TCL1A	UCP2	1.61	1.98E-134	0.509	0.256
B_c3_TCL1A	ARPC1B	1.49	4.23E-202	0.699	0.406
B_c3_TCL1A	LIMD2	1.30	7.94E-167	0.708	0.438
B_c3_TCL1A	HLA-DMB	1.25	4.80E-137	0.635	0.371
Plasma_c1_HSP	IGHA1	3.51	2.14E-188	0.621	0.305
Plasma_c1_HSP	JCHAIN	2.78	0	0.816	0.289
Plasma_c1_HSP	IGLC2	2.66	2.21E-39	0.543	0.357
Plasma_c1_HSP	ERLEC1	1.62	2.47E-271	0.585	0.173
Plasma_c1_HSP	SSR3	1.40	3.59E-255	0.665	0.274
Plasma_c2_TNFRSF17	JCHAIN	3.04	4.21E-287	0.938	0.35
Plasma_c2_TNFRSF17	IGLC2	2.50	4.15E-11	0.511	0.382
Plasma_c2_TNFRSF17	TNFRSF17	2.34	5.99E-81	0.388	0.129
Plasma_c2_TNFRSF17	IGHA1	2.23	4.56E-48	0.575	0.347
Plasma_c2_TNFRSF17	ERLEC1	1.60	1.43E-56	0.502	0.227
Plasmablast	RRM2	8.13	0	0.641	0.003
Plasmablast	MKI67	7.87	0	0.804	0.007
Plasmablast	CDK1	7.79	0	0.632	0.005
Plasmablast	TUBB	4.28	0	0.926	0.245
Plasmablast	TCL1A	2.30	8.13E-196	0.708	0.207

Supplementary Table 5 Top 5 differentially expressed genes across B and plasma cell states.

cluster	gene	avg_log2FC	p_val_adj	pct.1	pct.2
Mon_c1_CD14	FCN1	2.63	0E+00	0.68	0.199
Mon_c1_CD14	CD36	2.43	1E-98	0.377	0.194
Mon_c1_CD14	S100A12	2.16	3E-208	0.438	0.125
Mon_c1_CD14	S100A9	1.98	0E+00	0.923	0.584
Mon_c1_CD14	VCAN	1.93	3E-278	0.678	0.321
Mon_c2_CD16	CDKN1C	4.78	0E+00	0.808	0.109
Mon_c2_CD16	ZNF703	3.87	1E-299	0.532	0.069
Mon_c2_CD16	TCF7L2	3.74	0E+00	0.823	0.139
Mon_c2_CD16	FCN1	1.99	0E+00	0.872	0.207
Mon_c2_CD16	FGL2	1.50	1E-194	0.901	0.464
Mϕ_c1_IL1B	FCN1	3.68	0E+00	0.944	0.16
Mϕ_c1_IL1B	VCAN	3.21	0E+00	0.954	0.286
Mϕ_c1_IL1B	CD36	2.65	0E+00	0.701	0.162
Mϕ_c1_IL1B	S100A12	2.59	0E+00	0.746	0.089
Mϕ_c1_IL1B	S100A9	2.39	0E+00	0.992	0.566
Mϕ_c2_CCL18	CCL18	3.14	0E+00	0.485	0.084
Mϕ_c2_CCL18	GCHFR	2.58	0E+00	0.678	0.215
Mϕ_c2_CCL18	APOE	2.57	0E+00	0.954	0.387
Mϕ_c2_CCL18	MARCO	2.53	0E+00	0.536	0.151
Mϕ_c2_CCL18	IFI6	2.01	0E+00	0.805	0.373
Mϕ_c3_SLC40A1	SLC40A1	3.04	1.4e-318	0.451	0.116
Mϕ_c3_SLC40A1	GPR34	2.56	0E+00	0.558	0.168
Mϕ_c3_SLC40A1	C1QC	2.30	0E+00	0.847	0.287
Mϕ_c3_SLC40A1	SLCO2B1	2.23	0E+00	0.603	0.217
Mϕ_c3_SLC40A1	HLA-DRB5	1.52	9E-192	0.603	0.347
Mϕ_c4_VEGFA	SPP1	2.52	0E+00	0.822	0.385
Mϕ_c4_VEGFA	FN1	2.09	1E-260	0.528	0.218
Mϕ_c4_VEGFA	NDRG1	1.96	8E-250	0.522	0.225
Mϕ_c4_VEGFA	MARCO	1.76	0E+00	0.506	0.139
Mϕ_c4_VEGFA	VCAN	1.45	0E+00	0.702	0.262
Mϕ_c5_STMN1	TOP2A	6.90	0E+00	0.613	0.012
Mϕ_c5_STMN1	MKI67	6.82	0E+00	0.636	0.011
Mϕ_c5_STMN1	TYMS	6.47	0E+00	0.645	0.016
Mϕ_c5_STMN1	MZT2A	1.86	5E-270	0.64	0.243
Mϕ_c5_STMN1	CD1C	1.72	1E-79	0.219	0.064

Supplementary Table 6 Top 5 differentially expressed genes across monocyte and macrophage cell states.

cluster	gene	avg_log2FC	p_val_adj	pct.1	pct.2
DC_c1_CLEC9A	CLEC9A	6.84	1E-216	0.256	0.007
DC_c1_CLEC9A	CD1C	3.23	6E-187	0.463	0.065
DC_c1_CLEC9A	FCER1A	2.77	1E-117	0.48	0.12
DC_c1_CLEC9A	PPA1	2.30	1E-165	0.821	0.331
DC_c1_CLEC9A	CPVL	1.81	1E-37	0.648	0.405
DC_c2_CLEC10A	CD1C	5.06	0E+00	0.536	0.029
DC_c2_CLEC10A	FCER1A	4.09	0E+00	0.592	0.083
DC_c2_CLEC10A	CLEC10A	4.00	0E+00	0.62	0.088
DC_c2_CLEC10A	PPA1	2.53	0E+00	0.71	0.303
DC_c2_CLEC10A	HLA-DRB5	1.27	2E-159	0.618	0.382
DC_c3_LILRA4	SCT	10.24	0E+00	0.594	0.001
DC_c3_LILRA4	GZMB	9.82	0E+00	0.901	0.014
DC_c3_LILRA4	LRRC26	9.77	0E+00	0.579	0.001
DC_c3_LILRA4	LILRA4	8.16	0E+00	0.699	0.01
DC_c3_LILRA4	CD36	0.67	1E-23	0.415	0.198

Supplementary Table 7 Top 5 differentially expressed genes across dendritic cell states.

cluster	gene	avg_log2FC	p_val_adj	pct.1	pct.2
Mast_c1_CD52	CLU	4.11	0E+00	0.988	0.123
Mast_c1_CD52	RHEX	3.06	0E+00	0.518	0.058
Mast_c1_CD52	CSF1	1.61	6E-199	0.386	0.109
Mast_c1_CD52	FCER1A	1.05	2E-288	0.473	0.104
Mast_c1_CD52	CDKN1C	0.75	9E-46	0.224	0.111
Mast_c2_LMNA	RHEX	4.58	0E+00	0.646	0.031
Mast_c2_LMNA	CLU	3.15	0E+00	0.831	0.107
Mast_c2_LMNA	CSF1	3.14	0E+00	0.422	0.096
Mast_c2_LMNA	FCER1A	0.51	4E-92	0.284	0.108
Mast_c2_LMNA	VEGFA	0.47	5E-24	0.438	0.364
Neut_c1_CCL3	C15orf48	3.77	0E+00	0.738	0.389
Neut_c1_CCL3	CXCL8	3.37	2E-303	0.717	0.332
Neut_c1_CCL3	PHACTR1	2.68	5E-139	0.59	0.401
Neut_c1_CCL3	CCL3	2.39	1E-12	0.383	0.345
Neut_c1_CCL3	VEGFA	2.35	3E-34	0.428	0.369
Neut_c2_ISG15	FCGR3B	4.03	0E+00	0.907	0.09
Neut_c2_ISG15	CXCR2	3.66	0E+00	0.669	0.064
Neut_c2_ISG15	ISG15	3.08	1E-105	0.627	0.398
Neut_c2_ISG15	CSF3R	3.07	0E+00	0.9	0.356
Neut_c2_ISG15	IRF1	3.06	2E-177	0.703	0.405
Neut_c3_MYO1F	CSF3R	4.13	0E+00	0.919	0.353
Neut_c3_MYO1F	SLC25A37	3.75	0E+00	0.789	0.342
Neut_c3_MYO1F	MYO1F	3.70	7E-292	0.747	0.396
Neut_c3_MYO1F	LRRK2	3.45	1E-158	0.533	0.239
Neut_c3_MYO1F	IRF1	2.68	4E-57	0.498	0.408
Neut_c4_S100A9	CMTM2	5.66	0E+00	0.58	0.024
Neut_c4_S100A9	CXCR2	5.13	0E+00	0.553	0.031
Neut_c4_S100A9	FCGR3B	4.60	0E+00	0.769	0.045
Neut_c4_S100A9	CXCL8	3.55	2E-169	0.491	0.328
Neut_c4_S100A9	CSF3R	3.40	0E+00	0.796	0.326

Supplementary Table 8 Top 5 differentially expressed genes across mast cell and neutrophil states.

cluster	gene	avg_log2FC	p_val_adj	pct.1	pct.2
Fib_c1_MMP11	SDC1	4.98	0	0.685	0.05
Fib_c1_MMP11	COL11A1	4.76	0	0.836	0.109
Fib_c1_MMP11	MMP11	4.50	0	0.843	0.205
Fib_c1_MMP11	CTHRC1	3.11	0	0.974	0.435
Fib_c1_MMP11	FAP	1.59	0	0.821	0.386
Fib_c2_CXCL8	SAA1	7.78	0	0.593	0.041
Fib_c2_CXCL8	CXCL8	6.30	0	0.581	0.052
Fib_c2_CXCL8	ADM	4.04	0	0.739	0.183
Fib_c2_CXCL8	SERPINE1	2.85	0	0.826	0.358
Fib_c3_C7	PTGDS	2.99	0	0.858	0.329
Fib_c3_C7	COLEC11	2.93	0	0.607	0.119
Fib_c3_C7	C7	2.55	0	0.936	0.321
Fib_c3_C7	C3	1.77	0	0.872	0.349
Fib_c3_C7	PDGFRA	1.39	0	0.85	0.343
Fib_c4_CDH19	VIT	6.70	0	0.529	0.018
Fib_c4_CDH19	CDH19	3.56	0	0.516	0.033
Fib_c4_CDH19	CYP1B1	2.56	9.42E-207	0.723	0.333
Fib_c4_CDH19	IGF1	2.20	2.35E-121	0.539	0.229
Fib_c4_CDH19	PTGDS	0.73	3.13E-28	0.596	0.436
Fib_c5_HLA	CXCR4	5.23	0	0.666	0.139
Fib_c5_HLA	HLA-DRA	2.80	1.41E-239	0.715	0.369
Fib_c5_HLA	CXCL8	2.64	4.42E-66	0.21	0.063
Fib_c5_HLA	HLA-DRB1	1.71	9.00E-77	0.552	0.353
Fib_c5_HLA	CD74	1.24	3.21E-60	0.616	0.485
Fib_c6_IGF1	IGF1	3.69	0	0.707	0.175
Fib_c6_IGF1	TNXB	2.99	0	0.525	0.18
Fib_c6_IGF1	C3	2.27	0	0.852	0.407
Fib_c6_IGF1	PDGFRA	1.97	0	0.827	0.401
Fib_c6_IGF1	C7	1.74	0	0.863	0.396

Supplementary Table 9 Top 5 differentially expressed genes across fibroblast cell states.

cluster	gene	avg_log2FC	p_val_adj	pct.1	pct.2
EC_c1_ACKR1	ACKR1	6.60	0	0.779	0.038
EC_c1_ACKR1	SELP	5.61	0	0.717	0.023
EC_c1_ACKR1	CCL14	5.25	0	0.664	0.047
EC_c1_ACKR1	VWF	4.23	0	0.934	0.139
EC_c1_ACKR1	CD74	2.92	0	0.983	0.464
EC_c2_CCL21	TBX1	10.39	0	0.673	0.002
EC_c2_CCL21	CCL21	7.79	0	0.93	0.063
EC_c2_CCL21	FLT4	5.38	2.07e-315	0.729	0.046
EC_c2_CCL21	CAVIN2	4.32	2.23E-180	0.841	0.152
EC_c2_CCL21	CCL14	2.62	6.51E-41	0.341	0.074
EC_c3_RGCC	F2RL3	4.88	0	0.659	0.044
EC_c3_RGCC	EXOC3L2	4.43	0	0.357	0.026
EC_c3_RGCC	FLT1	3.97	0	0.806	0.087
EC_c3_RGCC	INHBB	3.78	0	0.421	0.043
EC_c3_RGCC	FLT4	3.73	1.12E-282	0.292	0.026
EC_c4_SEMA3G	SEMA3G	6.39	0	0.657	0.018
EC_c4_SEMA3G	GJA5	5.84	0	0.534	0.035
EC_c4_SEMA3G	JAG2	4.37	0	0.611	0.06
EC_c4_SEMA3G	INHBB	3.01	4.01E-235	0.361	0.065
EC_c4_SEMA3G	FLT1	2.99	0	0.787	0.126
EC_c5_STEAP4	EXOC3L2	2.62	0	0.518	0.045
EC_c5_STEAP4	F2RL3	2.56	0	0.731	0.084
EC_c5_STEAP4	CAVIN2	2.51	0	0.855	0.141
EC_c5_STEAP4	INHBB	2.46	0	0.574	0.066
EC_c5_STEAP4	FLT1	2.27	0	0.875	0.134

Supplementary Table 10 Top 5 differentially expressed genes across endothelial cell states.

Publication	Feature	Gene signatures
Elyada et al.[56]	CAFs	iCAF, myCAF, apCAF
Hwang et al.[61]	CAFs	Adhesive, Immunomodulatory, Myofibroblastic, Neurotropic
Schupp et al.[82]	Endothelial cells	Lymphatic, Arterial, Capillary, Venous
Bagaev et al.[203]	Macrophages	M1, M2, Phagocytosis
Zhou et al.[278]	Neutrophils	TAN, NET
Chu et al.[279]	Metabolism	OXPHOS, Glycolysis, Lipid metabolism, Fatty acid metabolism
Bagaev et al.[203]	TME	Angiogenesis, Antitumor cytokines, B cells, CAFs, Checkpoint molecules, Co-stimulatory ligands, Co-stimulatory receptors, Effector cell traffic, Effector cells, Endothelium, Granulocyte traffic, Immune Suppression by Myeloid Cells, Macrophage and DC traffic, Matrix, Matrix remodeling, MHCI, MHCI, Myeloid cell traffic, Neutrophil signature, Protumour cytokines, T cells, Th1 signature, Th2 signature, Treg, Treg and Th2 traffic, Tumour proliferation rate
Moffitt et al.[40]	Stromal	Activated Stroma, Normal Stroma
Moffitt et al.[40]	Malignant	Basal, Classical
Collisson et al.[39]	Malignant	Classical, Exocrine, QMA
Bailey et al.[41]	Malignant	Progenitor, ADEX, Immunogenic, Squamous
Hwang et al.[61]	Malignant	Acinar-like, Classical-like, Basaloid, Squamoid, Mesenchymal, Neuroendocrine, Neural-like progenitor
Gavish et al.[224]	Malignant	MP1-41

Supplementary Table 11 Curated gene signatures used in gene signature and pathway enrichment analyses.

Clinical characteristics		**Total**	**TCGA**	**CPTAC**	**ICGC**	**Puleo FRA**	**Puleo BEL**
		n=669	n=176	n=130	n=75	n=209	n=79
Gender	Female	299 (44.7%)	80 (45.5%)	61 (46.9%)	36 (48%)	85 (40.7%)	37 (46.8%)
	Male	370 (55.3%)	96 (54.5%)	69 (53.1%)	39 (52%)	124 (59.3%)	42 (53.2%)
Age	<65	179 (26.8%)	81 (46%)	65 (50%)	33 (44%)	0 (0%)	0 (0%)
	>=65	202 (30.2%)	95 (54%)	65 (50%)	42 (56%)	0 (0%)	0 (0%)
	no data	288 (43%)	0 (0%)	0 (0%)	0 (0%)	209 (100%)	79 (100%)
Stage	I	57 (8.5%)	21 (11.9%)	23 (17.7%)	5 (6.7%)	6 (2.9%)	2 (2.5%)
	II	334 (49.9%)	145 (82.4%)	55 (42.3%)	66 (88%)	47 (22.5%)	21 (26.6%)
	III	256 (38.3%)	4 (2.3%)	39 (30%)	1 (1.3%)	156 (74.6%)	56 (70.9%)
	IV	16 (2.4%)	4 (2.3%)	9 (6.9%)	3 (4%)	0 (0%)	0 (0%)
	no data	6 (0.9%)	2 (1.1%)	4 (3.1%)	0 (0%)	0 (0%)	0 (0%)

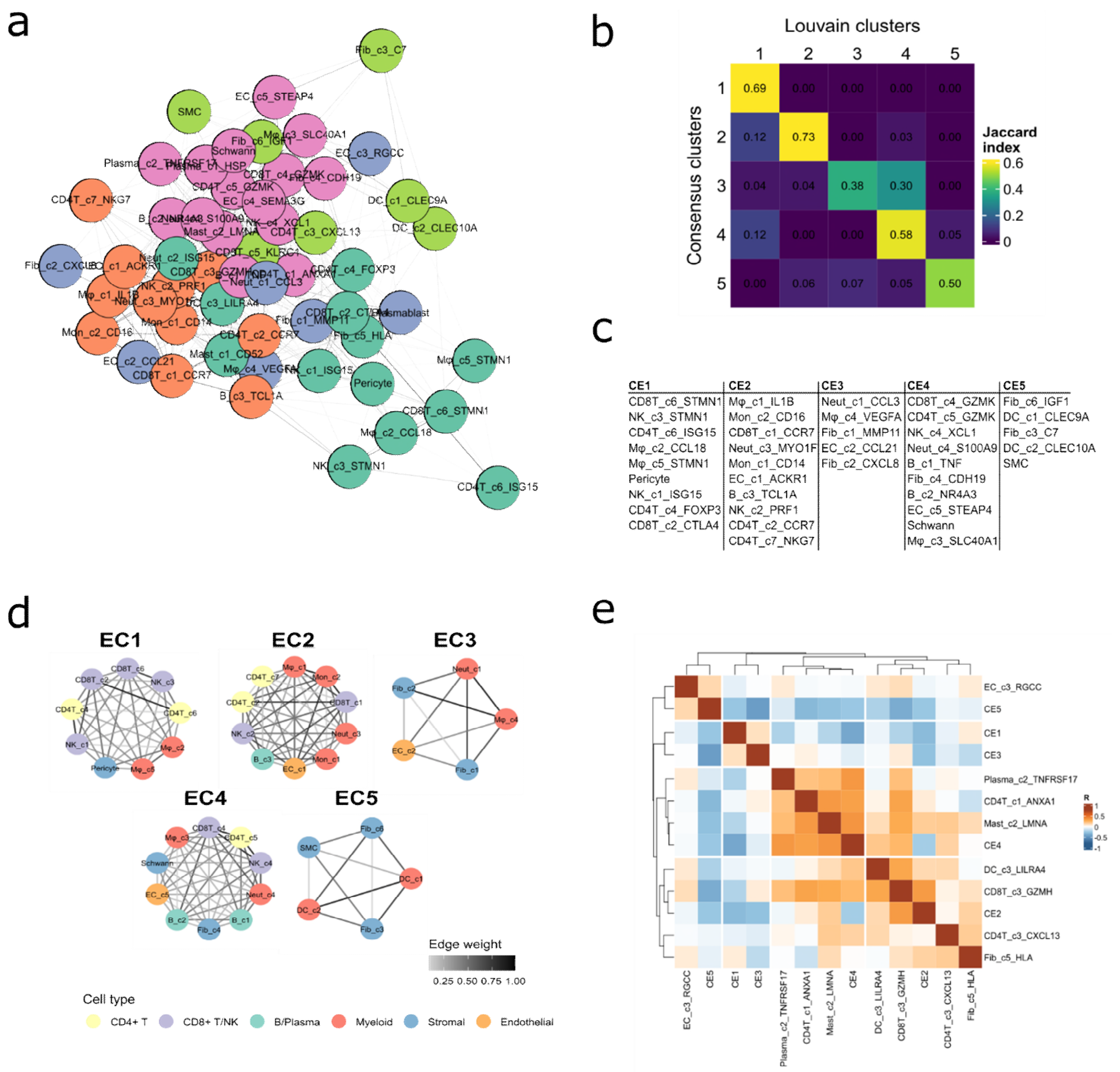
Supplementary Table 12 Clinical characteristics of PDAC bulk RNA-seq datasets.

Tumour	Non-malignant epithelial
Mucin 1 (MUC1)	Amylase 2A (AMY2A)
Cadherin 1 (CDH1)	AMY2B
KRT19	Protease serine 1 (PRSS1)
FXYD domain-containing ion transport regulator 3 (FXYD3)	Cystic Fibrosis Transmembrane Conductance Regulator (CFTR)
KRT17	Solute Carrier Family 4 Member 4 (SLC4A4)
KRT7	Bicaudal C homolog 1 (BICC1)
S100 calcium-binding protein A4 (S100A4)	
S100A6	
Trefoil factor 1 (TFF1)	
TFF3	

Supplementary Table 13 Tumour and non-malignant epithelial gene signatures. Derived from Zhou et al.[107]

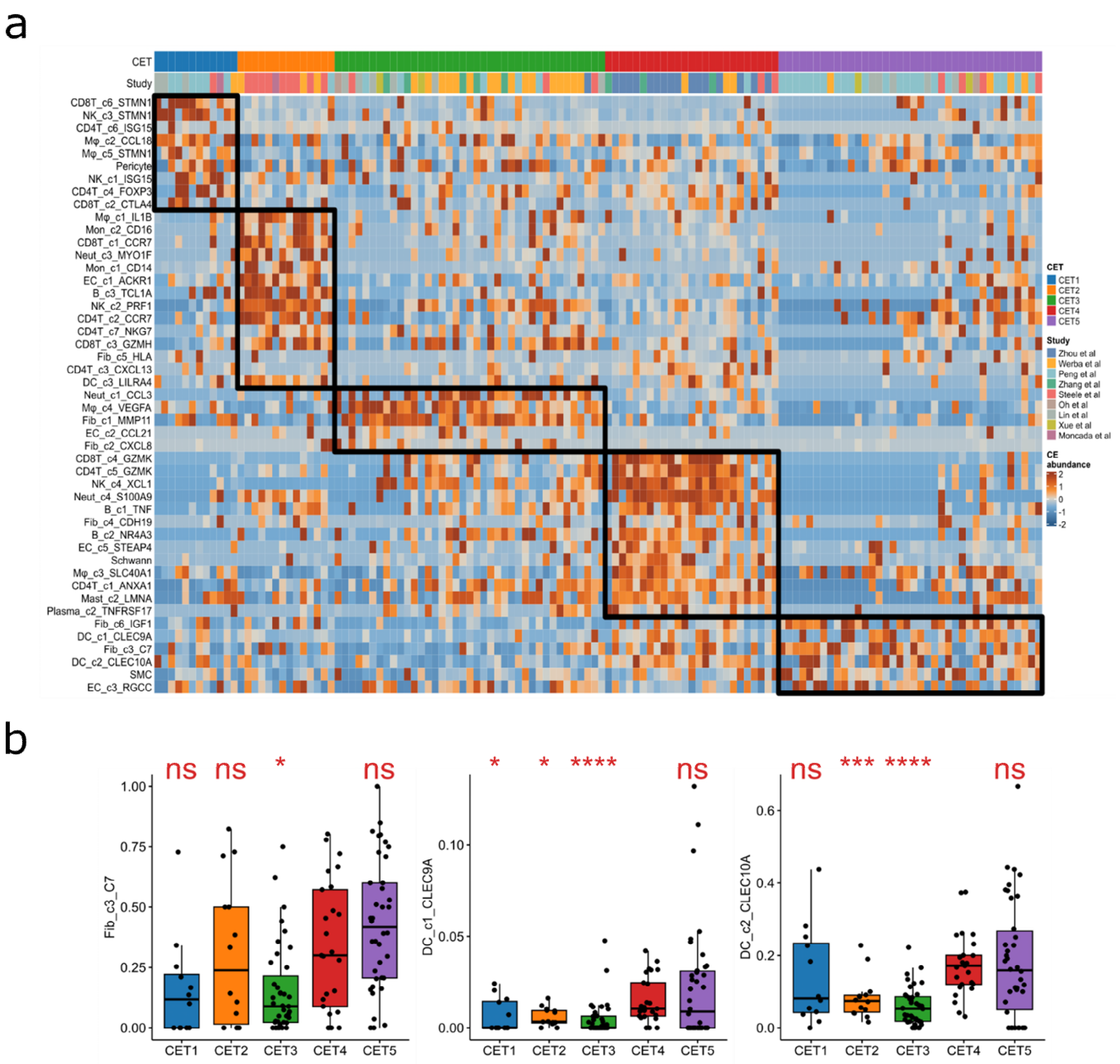
S1_Basal	S2_Classical	S3_Cycling_G2/M	S4_Progenitor	S5_Hypoxia	S6_Cycling_G1/S
CST6	REG4	MKI67	OLFM4	ADM	CDC6
TGFBI	TFF1	TOP2A	REG1A	HDAC9	GINS2
S100A9	MUC5AC	ASPM	CLDN2	TRIB3	CDC45
KRT6A	TFF2	CENPF	MUC6	DUSP5	E2F1
PTGS2	PHGR1	GTSE1	PIGR	DDIT3	ZNF367
SERPINE1	CDH17	HMMR	SERPINA1	GDF15	DSCC1
KRT16	CLDN18	UBE2C	CLU	CREB5	MCM6
THBS1	PSCA	DLGAP5	LYZ	PMAIP1	MCM2
FAM83A	CA9	CDKN3	NNMT	GADD45B	MCM5
KLK6	VSIG1	CENPE	FXYD2	SECTM1	BRCA1
PODXL	DUOXA2	KIF4A	PLEKHS1	GULP1	CHAF1A
S100A4	PLA2G10	NEK2	SERPINA3	PTPRR	EXO1
DCBLD2	SULT1C2	KIF23	CFTR	RGCC	HELLS
IGFBP6	CDHR2	TPX2	UGT2B15	RSAD2	UHRF1
LAMC2	CLRN3	CCNB2	APCS	SLC2A1	WDR76
S100A2	MUC13	KIF2C	LCN2	CITED2	CDC25A
CRABP2	PAQR8	CEP55	ADH1C	ERO1A	MCM4
IGFBP7	S100P	DEPDC1	COLCA1	IGFBP1	ORC6
LGALS1	MUC17	CDC20	CRISP3	IL23A	BRIP1
SERPINB2	MUC3A	NUSAP1	SCGB3A1	THSD7A	CCNE2
VCAN	AOC1	NUF2	BPIFB1	ANKRD37	CDT1
VIM	CDHR5	CCNB1	CXCL17	AREG	TEDC2
ALDH1A3	DUOX2	BIRC5	SOD3	C15orf48	XRCC2
FGFBP1	AGR2	NCAPG	CES1	CLDN23	ASF1B
FSTL1	CD68	PTTG1	COLGALT2	MUC4	BLM
WNT7A	HPGD	CENPA	CPVL	STEAP4	FEN1
FLNA	MYO1A	KIF14	FCGBP	IFIT2	POLA2
PI3	RAPGEFL1	PBK	IGFBP2	PIM1	TCF19
SNCG	CREB3L1	BUB1	KCNJ15	ZNF165	CCNE1
TMEM45A	IL1R2	KIF20A	MGP	ENO2	CHAF1B

Supplementary Table 14 Gene signatures of the identified cancer cell states.

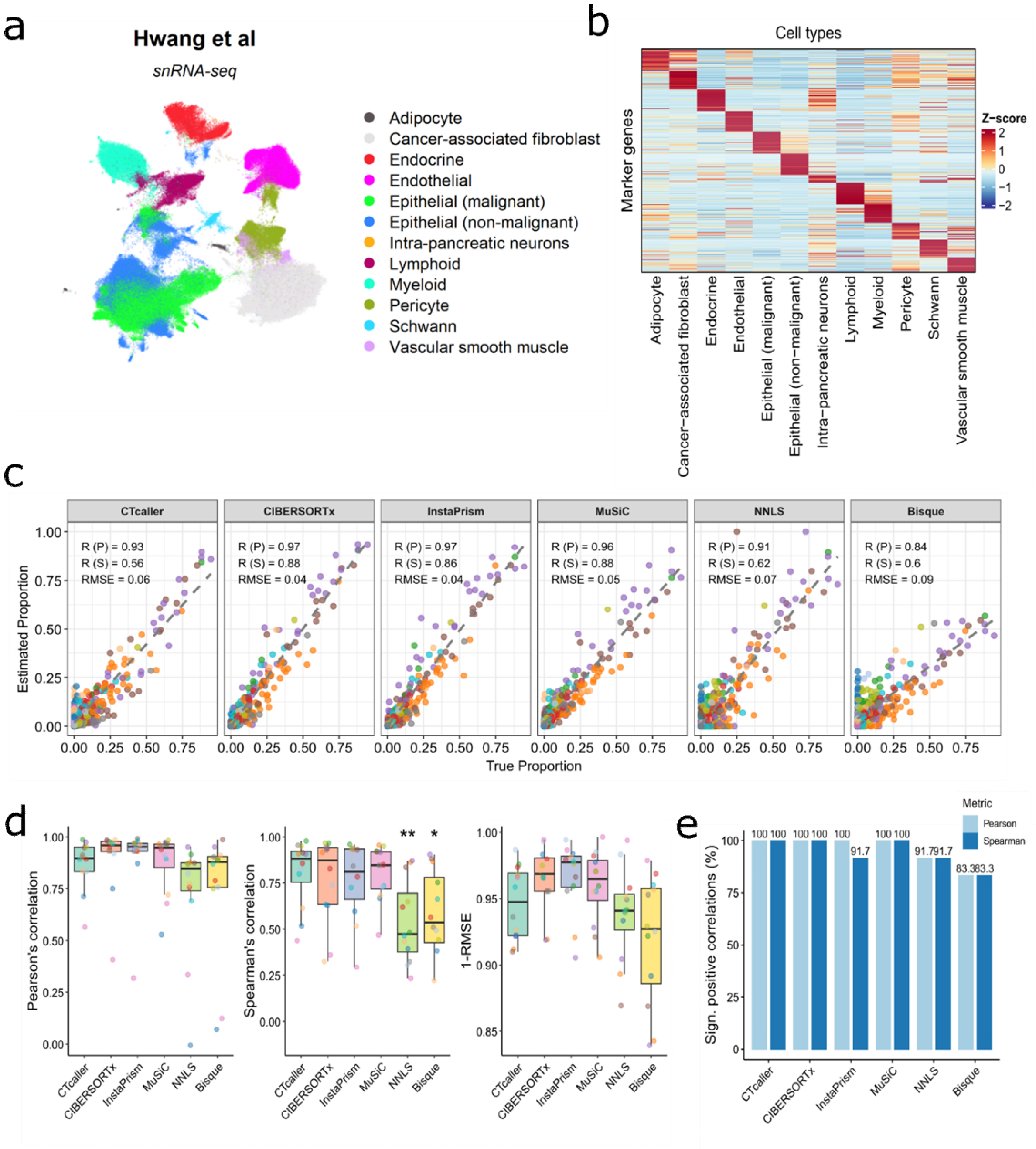


CE1	CE2	CE3	CE4	CE5
CD8T_c6_STMN1	Mφ_c1_IL1B	Neut_c1_CCL3	CD8T_c4_GZMK	Fib_c6_IGF1
NK_c3_STMN1	Mon_c2_CD16	Mφ_c4_VEGFA	CD4T_c5_GZMK	DC_c1_CLEC9A
CD4T_c6_ISG15	CD8T_c1_CCR7	Fib_c1_MMP11	NK_c4_XCL1	Fib_c3_C7
Mφ_c2_CCL18	Neut_c3_MYO1F	EC_c2_CCL21	Neut_c4_S100A9	DC_c2_CLEC10A
Mφ_c5_STMN1	Mon_c1_CD14	Fib_c2_CXCL8	B_c1_TNF	SMC
Pericyte	EC_c1_ACKR1		Fib_c4_CDH19	
NK_c1_ISG15	B_c3_TCL1A		B_c2_NR4A3	
CD4T_c4_FOXP3	NK_c2_PRF1		EC_c5_STEAP4	
CD8T_c2_CTLA4	CD4T_c2_CCR7		Schwann	
	CD4T_c7_NKG7		Mφ_c3_SLC40A1	

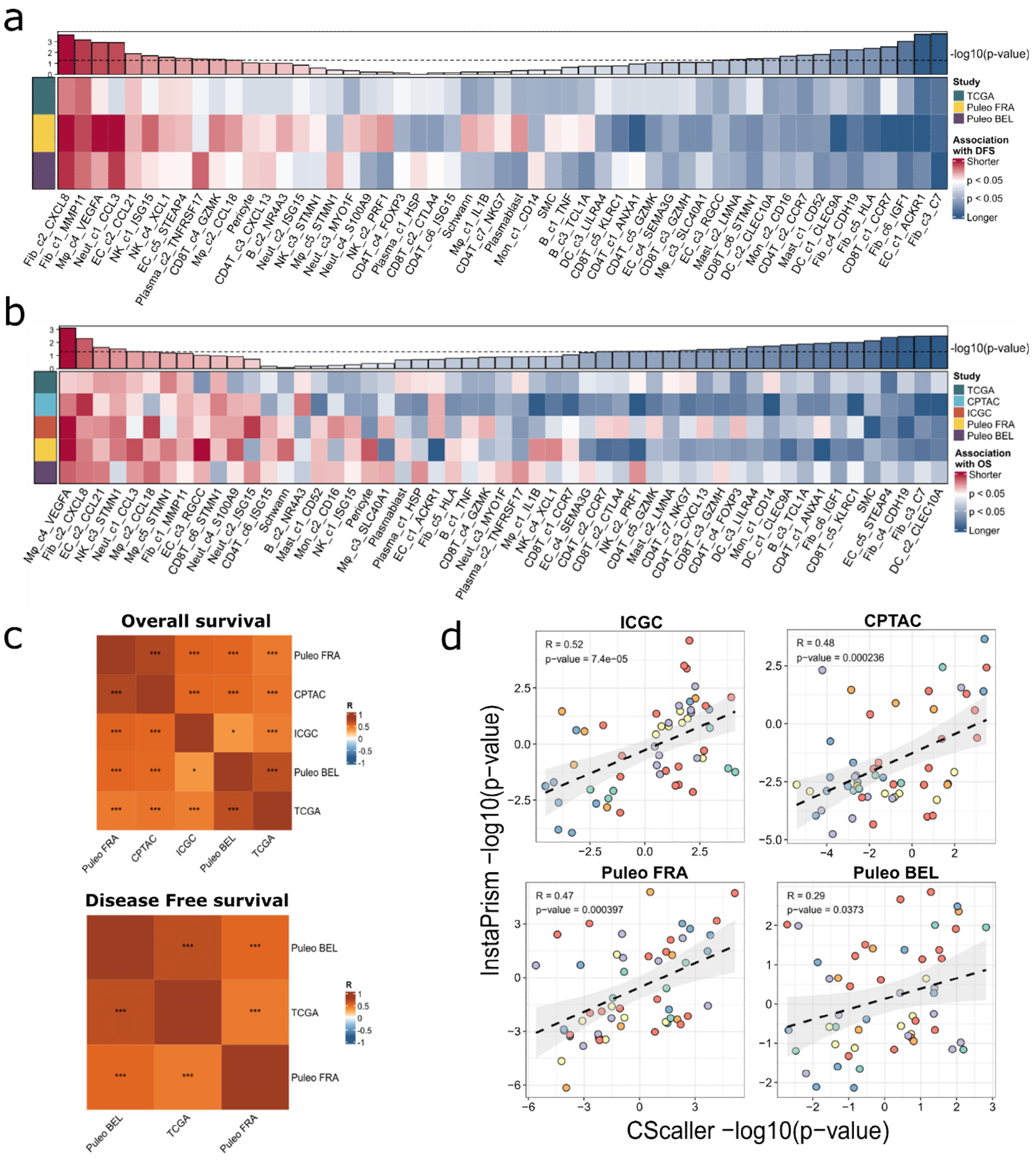
Supplementary Figure 1 Identification of PDAC cellular ecosystems. (a) Pearson correlation network of cell states clustered using the Louvain community algorithm. Nodes are coloured based on the Louvain clusters. Edge colour indicates the Pearson correlation coefficient. **(b)** Jaccard similarity matrix showing the overlap of cell states between the consensus and Louvain clusters. **(c)** Table showing the core cell states of CE1-5. **(d)** Network plots of the 5 core CEs, with nodes representing cell states labelled with short names and coloured by major cell type. Edge colour indicates the Pearson correlation coefficient. **(e)** Pearson correlation heatmap of remaining cell states and CE abundances.



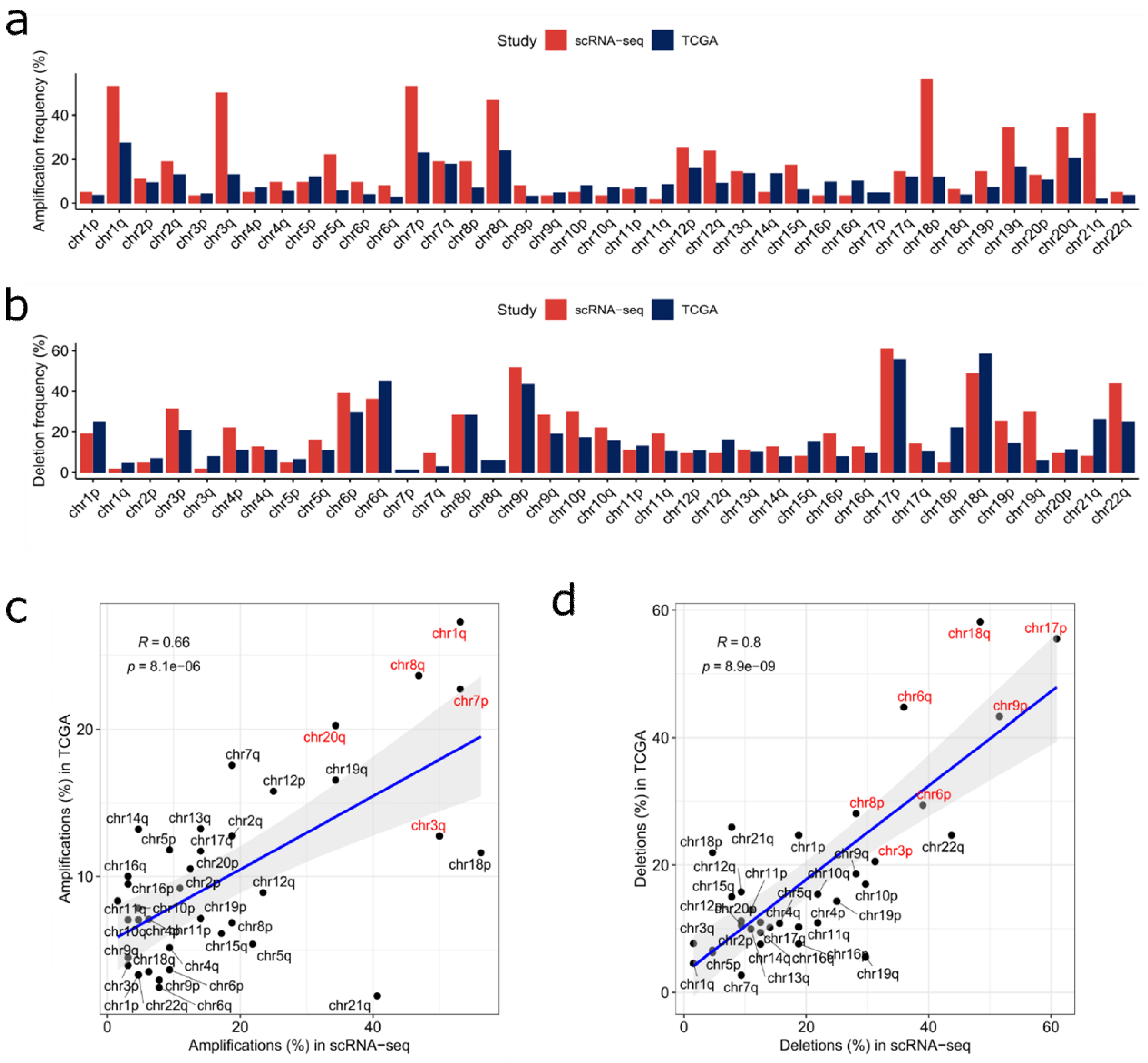
Supplementary Figure 2 Cell state abundance across individual CETs. (a) Heatmap showing the scaled cell state abundance across patient samples grouped by their corresponding CET. **(b)** Boxplots comparing the abundance of selected CE5 cell states between CET4 and the other CETs. Statistical significance was calculated using a two-sided Wilcoxon rank-sum test. *P<0.05, **P<0.01, ***P<0.001, ****P<0.0001.



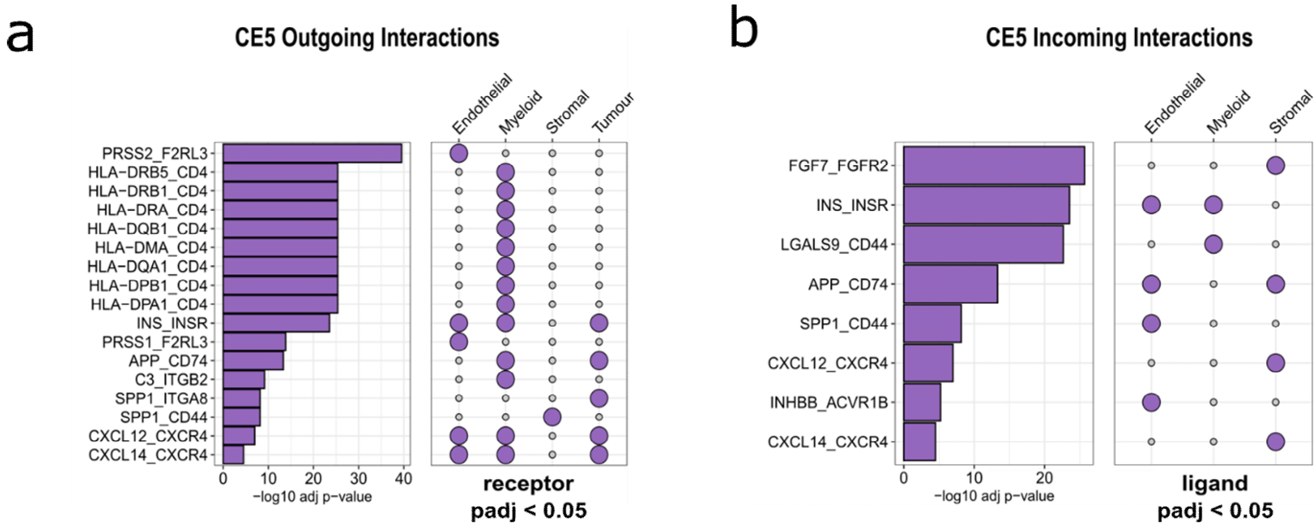
Supplementary Figure 3 Cell type deconvolution benchmarking in the snRNA-seq validation dataset. (a) UMAP embeddings of the snRNA-seq dataset (Hwang et al) coloured based on the pre-annotated cell type annotations. **(b)** Heatmap showing the refined cell type signature matrix used for deconvolution benchmarking. **(c)** Correlation of estimated cell type abundances with ground truth proportions facetted by deconvolution methods. Points are coloured based on the cell type. Both the Pearson's (P) and Spearman's (S) correlation coefficients are displayed along with the RMSE. **(d)** Boxplots comparing the cell type -based performance metrics across deconvolution methods. Statistical significance was calculated using a two-sided Wilcoxon rank-sum test using the CTcaller as reference. *P<0.05, **P<0.01, ***P<0.001. **(e)** Bar plots showing the frequency of cell types with significant positive correlations ($R > 0$ and $P < 0.05$) between their ground truth proportions and estimated proportions.



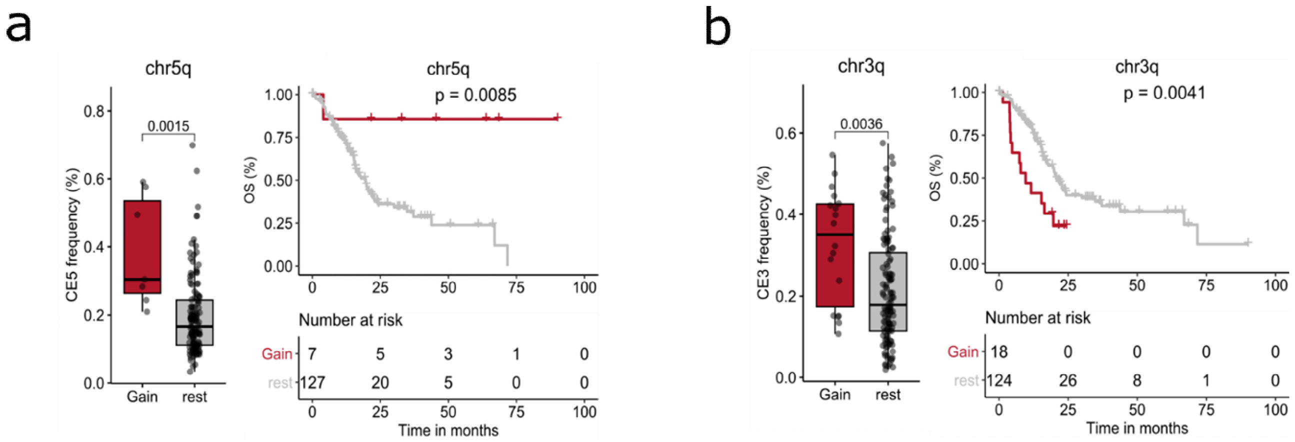
Supplementary Figure 4 Association of TME cell states with patient outcomes. (a) Heatmap showing the association of cell states with disease-free survival (DFS) through multivariate Cox regression modelling. If cell states are associated with worse outcomes (HR >= 1), the survival association is presented as -log10 (p-value). If cell states are associated with improved outcomes (HR < 1), the survival association is presented as -log10 (p-value) multiplied by -1. Bar plot on top represents the average the survival association for each cell state. **(b)** Heatmap showing the association of cell states with overall survival (DFS) using the InstaPrism deconvolution method. **(c)** Correlation matrix comparing the survival associations between cohorts for OS (top), and DFS (bottom). Statistical significance was calculated using Pearson's correlation. *P<0.05, **P<0.01, ***P<0.001. **(d)** Correlation scatter plots comparing the survival associations of cell states between the CScaller and InstaPrism deconvolution tools for each study. Points are coloured based on cell type. Statistical significance was calculated using Pearson's correlation.



Supplementary Figure 5 Comparison of copy number alteration frequency between the scRNA-seq and TCGA datasets. (a) Bar plot comparing the chromosome arm amplification frequency between the scRNA-seq atlas and TCGA. **(b)** Bar plot comparing the chromosome arm deletion frequency between the scRNA-seq atlas and TCGA. **(c)** Pearson correlation of the scRNA-seq and TCGA chromosome arm amplification frequencies. **(d)** Pearson correlation of the scRNA-seq and TCGA chromosome arm deletion frequencies. Frequent chromosome alterations reported in PDAC are coloured as red.



Supplementary Figure 6 Outgoing and Incoming communication signals of CE5. (a-b) Outgoing and incoming interactions between malignant cell states and TME cell types of CE5. The bar plot shows ligand-receptor pairs enriched in each CE, ranked by the -log10 (p-value), determined using a two-sided Wilcoxon rank-sum test. Dot plots show the enrichment of the receptor and ligand, for outgoing and incoming signals, respectively. Statistical significance was calculated using a two-sided Wilcoxon rank-sum test.



Supplementary Figure 7 Association of broad CNAs with CE frequencies. (a) Association of CE5 frequency and chr5q amplification. Kaplan-Meier OS curve of the amplified chr5q. Statistical significance was determined using the two-sided log-rank test. **(b)** Association of CE3 frequency and chr3q amplification. Kaplan-Meier OS curve of the amplified chr3q. Statistical significance was determined using the two-sided log-rank test.

References

1. Leiphrakpam PD, Chowdhury S, Zhang M, Bajaj V, Dhir M, Are C. Trends in the Global Incidence of Pancreatic Cancer and a Brief Review of its Histologic and Molecular Subtypes. J Gastrointest Canc 2025;56(1):71.

2. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2024;74(3):229–63.

3. Park W, Chawla A, O'Reilly EM. Pancreatic Cancer: A Review. JAMA 2021;326(9):851–62.

4. Klein AP. Pancreatic cancer epidemiology: understanding the role of lifestyle and inherited risk factors. Nat Rev Gastroenterol Hepatol 2021;18(7):493–502.

5. Groot VP, Rezaee N, Wu W, Cameron JL, Fishman EK, Hruban RH et al. Patterns, Timing, and Predictors of Recurrence Following Pancreatectomy for Pancreatic Ductal Adenocarcinoma. Annals of Surgery 2018;267(5):936.

6. Oettle H, Neuhaus P, Hochhaus A, Hartmann JT, Gellert K, Ridwelski K et al. Adjuvant chemotherapy with gemcitabine and long-term outcomes among patients with resected pancreatic cancer: the CONKO-001 randomized trial. JAMA 2013;310(14):1473–81.

7. Neoptolemos JP, Palmer DH, Ghaneh P, Psarelli EE, Valle JW, Halloran CM et al. Comparison of adjuvant gemcitabine and capecitabine with gemcitabine monotherapy in patients with resected pancreatic cancer (ESPAC-4): a multicentre, open-label, randomised, phase 3 trial. The Lancet 2017;389(10073):1011–24.

8. Kung H, Yu J. Targeted therapy for pancreatic ductal adenocarcinoma: Mechanisms and clinical study. MedComm (2020) 2023;4(2):e216.

9. Conroy T, Hammel P, Hebbar M, Ben Abdelghani M, Wei AC, Raoul J et al. FOLFIRINOX or Gemcitabine as Adjuvant Therapy for Pancreatic Cancer. N Engl J Med 2018;379(25):2395–406.

10. Janssen QP, O'Reilly EM, van Eijck CHJ, Groot Koerkamp B. Neoadjuvant Treatment in Patients With Resectable and Borderline Resectable Pancreatic Cancer. Front Oncol 2020;10:41.

11. Nagakawa Y, Sahara Y, Hosokawa Y, Murakami Y, Yamaue H, Satoi S et al. Clinical Impact of Neoadjuvant Chemotherapy and Chemoradiotherapy in Borderline Resectable Pancreatic Cancer: Analysis of 884 Patients at Facilities Specializing in Pancreatic Surgery. Ann Surg Oncol 2019;26(6):1629–36.

12. Jang J, Han Y, Lee H, Kim S, Kwon W, Lee K et al. Oncological Benefits of Neoadjuvant Chemoradiation With Gemcitabine Versus Upfront Surgery in Patients With Borderline Resectable Pancreatic Cancer: A Prospective, Randomized, Open-label, Multicenter Phase 2/3 Trial. Annals of Surgery 2018;268(2):215.

13. Versteijne E, van Dam JL, Suker M, Janssen QP, Groothuis K, Akkermans-Vogelaar JM et al. Neoadjuvant Chemoradiotherapy Versus Upfront Surgery for Resectable and Borderline Resectable Pancreatic Cancer: Long-Term Results of the Dutch Randomized PREOPANC Trial. J Clin Oncol 2022;40(11):1220–30.

14. Conroy T, Desseigne F, Ychou M, Bouché O, Guimbaud R, Bécouarn Y et al. FOLFIRINOX versus Gemcitabine for Metastatic Pancreatic Cancer. New England Journal of Medicine 2011;364(19):1817–25.

15. Dusetti N, Bachet JB, Chanez B, Neuzillet C, de Mestier L, Williet N et al. Medical management of pancreatic cancer: from personalization to broadening treatment strategies. Cancer Treatment Reviews 2025;138:102973.

16. Golan T, Hammel P, Reni M, Van Cutsem E, Macarulla T, Hall MJ et al. Maintenance Olaparib for Germline BRCA-Mutated Metastatic Pancreatic Cancer. N Engl J Med 2019;381(4):317–27.

17. Momtaz P, O'Connor CA, Chou JF, Capanu M, Park W, Bandlamudi C et al. Pancreas cancer and BRCA: A critical subset of patients with improving therapeutic outcomes. Cancer 2021;127(23):4393–402.

18. Brown TJ, Reiss KA. PARP Inhibitors in Pancreatic Cancer. Cancer J 2021;27(6):465–75.

19. Balsano R, Zanuso V, Pirozzi A, Rimassa L, Bozzarelli S. Pancreatic Ductal Adenocarcinoma and Immune Checkpoint Inhibitors: The Gray Curtain of Immunotherapy and Spikes of Lights. Current Oncology 2023;30(4):3871–85.

20. Royal RE, Levy C, Turner K, Mathur A, Hughes M, Kammula US et al. Phase 2 Trial of Single Agent Ipilimumab (Anti-CTLA-4) for Locally Advanced or Metastatic Pancreatic Adenocarcinoma. Journal of Immunotherapy 2010;33(8):828.

21. O'Reilly EM, Oh D, Dhani N, Renouf DJ, Lee MA, Sun W et al. Durvalumab With or Without Tremelimumab for Patients With Metastatic Pancreatic Ductal Adenocarcinoma: A Phase 2 Randomized Clinical Trial. JAMA Oncol 2019;5(10):1431–8.

22. Kamath SD, Kalyan A, Kircher S, Nimeiri H, Fought AJ, Benson A et al. Ipilimumab and Gemcitabine for Advanced Pancreatic Cancer: A Phase Ib Study. Oncologist 2020;25(5):e808–15.

23. Marabelle A, O'Malley DM, Hendifar AE, Ascierto PA, Motola-Kuba D, Penel N et al. Pembrolizumab in microsatellite-instability-high and mismatch-repair-deficient advanced solid tumors: updated results of the KEYNOTE-158 trial. Nat Cancer 2025;6(2):253–8.

24. Humphris JL, Patch A, Nones K, Bailey PJ, Johns AL, McKay S et al. Hypermutation In Pancreatic Cancer. Gastroenterology 2017;152(1):68,74.e2.

25. Zugasti I, Espinosa-Aroca L, Fidyt K, Mulens-Arias V, Diaz-Beya M, Juan M et al. CAR-T cell therapy for cancer: current challenges and future directions. Sig Transduct Target Ther 2025;10(1):210.

26. Beatty GL, O'Hara MH, Lacey SF, Torigian DA, Nazimuddin F, Chen F et al. Activity of Mesothelin-Specific Chimeric Antigen Receptor T Cells Against Pancreatic Carcinoma Metastases in a Phase 1 Trial. Gastroenterology 2018;155(1):29–32.

27. Haas AR, Tanyi JL, O'Hara MH, Gladney WL, Lacey SF, Torigian DA et al. Phase I Study of Lentiviral-Transduced Chimeric Antigen Receptor-Modified T Cells Recognizing Mesothelin in Advanced Solid Cancers. Molecular Therapy 2019;27(11):1919–29.

28. Yang J, Lin P, Yang M, Liu W, Fu X, Liu D et al. Integrated genomic and transcriptomic analysis reveals unique characteristics of hepatic metastases and pro-metastatic role of complement C1q in pancreatic ductal adenocarcinoma. Genome Biology 2021;22(1):4.

29. Waddell N, Pajic M, Patch A, Chang DK, Kassahn KS, Bailey P et al. Whole genomes redefine the mutational landscape of pancreatic cancer. Nature 2015;518(7540):495–501.

30. Biankin AV, Waddell N, Kassahn KS, Gingras M, Muthuswamy LB, Johns AL et al. Pancreatic cancer genomes reveal aberrations in axon guidance pathway genes. Nature 2012;491(7424):399–405.

31. Connor AA, Gallinger S. Pancreatic cancer evolution and heterogeneity: integrating omics and clinical data. Nat Rev Cancer 2022;22(3):131–42.

32. Notta F, Chan-Seng-Yue M, Lemire M, Li Y, Wilson GW, Connor AA et al. A renewed model of pancreatic cancer evolution based on genomic rearrangement patterns. Nature 2016;538(7625):378–82.

33. Al-Sukhni W, Borgida A, Rothenmund H, Holter S, Semotiuk K, Grant R et al. Screening for pancreatic cancer in a high-risk cohort: an eight-year experience. J Gastrointest Surg 2012;16(4):771–83.

34. Bachet JB, Maréchal R, Demetter P, Bonnetain F, Couvelard A, Svrcek M et al. Contribution of CXCR4 and SMAD4 in predicting disease progression pattern and benefit from adjuvant chemotherapy in resected pancreatic adenocarcinoma. Annals of Oncology 2012;23(9):2327–35.

35. Varghese AM, Perry MA, Chou JF, Nandakumar S, Muldoon D, Erakky A et al. Clinicogenomic landscape of pancreatic adenocarcinoma identifies KRAS mutant dosage as prognostic of overall survival. Nat Med 2025;31(2):466.

36. Xu X, Chen X, Xu R, Huo Z, Li C, Nowsheen S et al. Combination of PARP and KRASG12D inhibitors enhances therapeutic efficacy by exploiting vulnerabilities in PDAC. Nat Commun 2026;17(1):3118.

37. Becker JH, Metropulos AE, Spaulding C, Marinelarena AM, Shields MA, Principe DR et al. Targeting BCL2 with Venetoclax Enhances the Efficacy of the KRAS G12D Inhibitor MRTX1133 in Pancreatic Cancer. Cancer Res 2024;84(21):3629.

38. Qin S, Li J, Bai Y, Wang Z, Chen Z, Xu R et al. Nimotuzumab Plus Gemcitabine for K-Ras Wild-Type Locally Advanced or Metastatic Pancreatic Cancer. JCO 2023;41(33):5163.

39. Collisson EA, Sadanandam A, Olson P, Gibb WJ, Truitt M, Gu S et al. Subtypes of pancreatic ductal adenocarcinoma and their differing responses to therapy. Nat Med 2011;17(4):500–3.

40. Moffitt RA, Marayati R, Flate EL, Volmar KE, Loeza SGH, Hoadley KA et al. Virtual microdissection identifies distinct tumor- and stroma-specific subtypes of pancreatic ductal adenocarcinoma. Nat Genet 2015;47(10):1168–78.

41. Bailey P, Chang DK, Nones K, Johns AL, Patch A, Gingras M et al. Genomic analyses identify molecular subtypes of pancreatic cancer. Nature 2016;531(7592):47–52.

42. Chan-Seng-Yue M, Kim JC, Wilson GW, Ng K, Figueroa EF, O'Kane GM et al. Transcription phenotypes of pancreatic cancer are driven by genomic events during tumor evolution. Nat Genet 2020;52(2):231–40.

43. Mueller S, Engleitner T, Maresch R, Zukowska M, Lange S, Kaltenbacher T et al. Evolutionary routes and KRAS dosage define pancreatic cancer phenotypes. Nature 2018;554(7690):62–8.

44. Eyres M, Lanfredini S, Xu H, Burns A, Blake A, Willenbrock F et al. TET2 Drives 5hmc Marking of *GATA6* and Epigenetically Defines Pancreatic Ductal Adenocarcinoma Transcriptional Subtypes. Gastroenterology 2021;161(2):653,668.e16.

45. Tu M, Klein L, Espinet E, Georgomanolis T, Wegwitz F, Li X et al. TNF-α-producing macrophages determine subtype identity and prognosis via AP1 enhancer reprogramming in pancreatic cancer. Nat Cancer 2021;2(11):1185–203.

46. Aung KL, Fischer SE, Denroche RE, Jang G, Dodd A, Creighton S et al. Genomics-Driven Precision Medicine for Advanced Pancreatic Cancer: Early Results from the COMPASS Trial. Clin Cancer Res 2018;24(6):1344–54.

47. Aung KL, Fischer SE, Denroche RE, Jang G, Dodd A, Creighton S et al. Genomics-Driven Precision Medicine for Advanced Pancreatic Cancer: Early Results from the COMPASS Trial. Clin Cancer Res 2018;24(6):1344–54.

48. Roa-Peña L, Leiton CV, Babu S, Pan C, Vanner EA, Akalin A et al. Keratin 17 identifies the most lethal molecular subtype of pancreatic cancer. Sci Rep 2019;9(1):11239.

49. Zhou X, An J, Kurilov R, Brors B, Hu K, Peccerella T et al. Persister cell phenotypes contribute to poor patient outcomes after neoadjuvant chemotherapy in PDAC. Nat Cancer 2023;4(9):1362–81.

50. Miyabayashi K, Baker LA, Deschênes A, Traub B, Caligiuri G, Plenker D et al. Intraductal Transplantation Models of Human Pancreatic Ductal Adenocarcinoma Reveal Progressive Transition of Molecular Subtypes. Cancer Discov 2020;10(10):1566–89.

51. Hwang RF, Moore T, Arumugam T, Ramachandran V, Amos KD, Rivera A et al. Cancer-Associated Stromal Fibroblasts Promote Pancreatic Tumor Progression. Cancer Res 2008;68(3):918–26.

52. Özdemir BC, Pentcheva-Hoang T, Carstens JL, Zheng X, Wu C, Simpson TR et al. Depletion of carcinoma-associated fibroblasts and fibrosis induces immunosuppression and accelerates pancreas cancer with reduced survival. Cancer Cell 2014;25(6):719–34.

53. Rhim AD, Oberstein PE, Thomas DH, Mirek ET, Palermo CF, Sastra SA et al. Stromal elements act to restrain, rather than support, pancreatic ductal adenocarcinoma. Cancer Cell 2014;25(6):735–47.

54. Sahai E, Astsaturov I, Cukierman E, DeNardo DG, Egeblad M, Evans RM et al. A framework for advancing our understanding of cancer-associated fibroblasts. Nat Rev Cancer 2020;20(3):174–86.

55. Öhlund D, Handly-Santana A, Biffi G, Elyada E, Almeida AS, Ponz-Sarvise M et al. Distinct populations of inflammatory fibroblasts and myofibroblasts in pancreatic cancer. J Exp Med 2017;214(3):579–96.

56. Elyada E, Bolisetty M, Laise P, Flynn WF, Courtois ET, Burkhart RA et al. Cross-Species Single-Cell Analysis of Pancreatic Ductal Adenocarcinoma Reveals Antigen-Presenting Cancer-Associated Fibroblasts. Cancer Discov 2019;9(8):1102–23.

57. Feig C, Jones JO, Kraman M, Wells RJB, Deonarine A, Chan DS et al. Targeting CXCL12 from FAP-expressing carcinoma-associated fibroblasts synergizes with anti-PD-L1 immunotherapy in pancreatic cancer. Proc Natl Acad Sci U S A 2013;110(50):20212–7.

58. Hallett RM, Bonfill-Teixidor E, Iurlaro R, Arias A, Raman S, Bayliss P et al. Therapeutic Targeting of LIF Overcomes Macrophage-mediated Immunosuppression of the Local Tumor Microenvironment. Clin Cancer Res 2023;29(4):791–804.

59. Yang X, Lin Y, Shi Y, Li B, Liu W, Yin W et al. FAP Promotes Immunosuppression by Cancer-Associated Fibroblasts in the Tumor Microenvironment via STAT3–CCL2 Signaling. Cancer Res 2016;76(14):4124–35.

60. Huang H, Wang Z, Zhang Y, Pradhan RN, Ganguly D, Chandra R et al. Mesothelial cell-derived antigen-presenting cancer-associated fibroblasts induce expansion of regulatory T cells in pancreatic cancer. Cancer Cell 2022;40(6):656,673.e7.

61. Hwang WL, Jagadeesh KA, Guo JA, Hoffman HI, Yadollahpour P, Reeves JW et al. Single-nucleus and spatial transcriptome profiling of pancreatic cancer identifies multicellular dynamics associated with neoadjuvant treatment. Nat Genet 2022;54(8):1178–91.

62. Chen K, Wang Q, Li M, Guo H, Liu W, Wang F et al. Single-cell RNA-seq reveals dynamic change in tumor microenvironment during pancreatic ductal adenocarcinoma malignant progression. eBioMedicine 2021;66

63. Lavie D, Ben-Shmuel A, Erez N, Scherz-Shouval R. Cancer-associated fibroblasts in the single-cell era. Nat Cancer 2022;3(7):793–807.

64. Grünwald BT, Devisme A, Andrieux G, Vyas F, Aliar K, McCloskey CW et al. Spatially confined sub-tumor microenvironments in pancreatic cancer. Cell 2021;184(22):5577,5592.e18.

65. Maurer C, Holmstrom SR, He J, Laise P, Su T, Ahmed A et al. Experimental microdissection enables functional harmonisation of pancreatic cancer subtypes. Gut 2019;68(6):1034–43.

66. Falcomatà C, Bärthel S, Schneider G, Rad R, Schmidt-Supprian M, Saur D. Context-Specific Determinants of the Immunosuppressive Tumor Microenvironment in Pancreatic Cancer. Cancer Discov 2023;13(2):278.

67. Marabelle A, Le DT, Ascierto PA, Di Giacomo AM, De Jesus-Acosta A, Delord J et al. Efficacy of Pembrolizumab in Patients With Noncolorectal High Microsatellite Instability/Mismatch Repair–Deficient Cancer: Results From the Phase II KEYNOTE-158 Study. J Clin Oncol 2020;38(1):1–10.

68. Steele NG, Carpenter ES, Kemp SB, Sirihorachai VR, The S, Delrosario L et al. Multimodal mapping of the tumor and peripheral blood immune landscape in human pancreatic cancer. Nat Cancer 2020;1(11):1097–112.

69. Schalck A, Sakellariou-Thompson D, Forget M, Sei E, Hughes TG, Reuben A et al. Single-Cell Sequencing Reveals Trajectory of Tumor-Infiltrating Lymphocyte States in Pancreatic Cancer. Cancer Discov 2022;12(10):2330–49.

70. Freed-Pastor WA, Lambert LJ, Ely ZA, Pattada NB, Bhutkar A, Eng G et al. The CD155/TIGIT axis promotes and maintains immune evasion in neoantigen-expressing pancreatic cancer. Cancer Cell 2021;39(10):1342,1360.e14.

71. Avşar G, Pir P. An integrated study to decipher immunosuppressive cellular communication in the PDAC environment. npj Syst Biol Appl 2023;9(1):56.

72. Jainarayanan A, Mouroug-Anand N, Arbe-Barnes EH, Bush AJ, Bashford-Rogers R, Frampton A et al. Pseudotime dynamics of T cells in pancreatic ductal adenocarcinoma inform distinct functional states within the regulatory and cytotoxic T cells. iScience 2023;26(4)

73. Binnewies M, Pollack JL, Rudolph J, Dash S, Abushawish M, Lee T et al. Targeting TREM2 on tumor-associated macrophages enhances immunotherapy. Cell Reports 2021;37(3)

74. Kemp SB, Carpenter ES, Steele NG, Donahue KL, Nwosu ZC, Pacheco A et al. Apolipoprotein E Promotes Immune Suppression in Pancreatic Cancer through NF-κB–Mediated Production of CXCL1. Cancer Res 2021;81(16):4305.

75. Caronni N, La Terza F, Vittoria FM, Barbiera G, Mezzanzanica L, Cuzzola V et al. IL-1β+ macrophages fuel pathogenic inflammation in pancreatic cancer. Nature 2023;623(7986):415–22.

76. Candido JB, Morton JP, Bailey P, Campbell AD, Karim SA, Jamieson T et al. CSF1R+ Macrophages Sustain Pancreatic Tumor Growth through T Cell Suppression and Maintenance of Key Gene Programs that Define the Squamous Subtype. Cell Reports 2018;23(5):1448–60.

77. Tu M, Klein L, Espinet E, Georgomanolis T, Wegwitz F, Li X et al. TNF-α-producing macrophages determine subtype identity and prognosis via AP1 enhancer reprogramming in pancreatic cancer. Nat Cancer 2021;2(11):1185–203.

78. Somerville TD, Biffi G, Daßler-Plenker J, Hur SK, He X, Vance KE et al. Squamous trans-differentiation of pancreatic cancer cells promotes stromal inflammation. eLife 2020;9:e53381.

79. Raghavan S, Winter PS, Navia AW, Williams HL, Denadel A, Lowder KE et al. Microenvironment drives cell state, plasticity, and drug response in pancreatic cancer. Cell 2021;184(25):6119.

80. Oh K, Yoo YJ, Torre-Healy LA, Rao M, Fassler D, Wang P et al. Coordinated single-cell tumor microenvironment dynamics reinforce pancreatic cancer subtype. Nat Commun 2023;14(1):5226.

81. Steuernagel L, Lam BYH, Klemm P, Dowsett GKC, Bauder CA, Tadross JA et al. HypoMap—a unified single-cell gene expression atlas of the murine hypothalamus. Nat Metab 2022;4(10):1402–19.

82. Schupp JC, Adams TS, Cosme C, Raredon MSB, Yuan Y, Omote N et al. Integrated Single-Cell Atlas of Endothelial Cells of the Human Lung. Circulation 2021;144(4):286.

83. Buechler MB, Pradhan RN, Krishnamurty AT, Cox C, Calviello AK, Wang AW et al. Cross-tissue organization of the fibroblast lineage. Nature 2021;593(7860):575–9.

84. Salcher S, Sturm G, Horvath L, Untergasser G, Kuempers C, Fotakis G et al. High-resolution single-cell atlas reveals diversity and plasticity of tissue-resident neutrophils in non-small cell lung cancer. Cancer Cell 2022;40(12):1503,1520.e8.

85. Shi Q, Chen Y, Li Y, Qin S, Yang Y, Gao Y et al. Cross-tissue multicellular coordination and its rewiring in cancer. Nature 2025;643(8071):529.

86. Zhang S, Fang W, Zhou S, Zhu D, Chen R, Gao X et al. Single cell transcriptomic analyses implicate an immunosuppressive tumor microenvironment in pancreatic cancer liver metastasis. Nat Commun 2023;14(1):1–19.

87. Werba G, Weissinger D, Kawaler EA, Zhao E, Kalfakakou D, Dhara S et al. Single-cell RNA sequencing reveals the effects of chemotherapy on human pancreatic adenocarcinoma and its tumor microenvironment. Nat Commun 2023;14(1):1–16.

88. Cui Zhou D, Jayasinghe RG, Chen S, Herndon JM, Iglesia MD, Navale P et al. Spatially restricted drivers and transitional cell populations cooperate with the microenvironment in untreated and chemo-resistant pancreatic cancer. Nat Genet 2022;54(9):1390–405.

89. Lin W, Noel P, Borazanci EH, Lee J, Amini A, Han IW et al. Single-cell transcriptome analysis of tumor and stromal compartments of pancreatic ductal adenocarcinoma primary tumors and metastatic lesions. Genome Med 2020;12(1):80.

90. Moncada R, Barkley D, Wagner F, Chiodin M, Devlin JC, Baron M et al. Integrating microarray-based spatial transcriptomics and single-cell RNA-seq reveals tissue architecture in pancreatic ductal adenocarcinomas. Nat Biotechnol 2020;38(3):333–42.

91. Oh K, Yoo YJ, Torre-Healy LA, Rao M, Fassler D, Wang P et al. Coordinated single-cell tumor microenvironment dynamics reinforce pancreatic cancer subtype. Nat Commun 2023;14(1):1–13.

92. Peng J, Sun B, Chen C, Zhou J, Chen Y, Chen H et al. Single-cell RNA-seq highlights intra-tumoral heterogeneity and malignant progression in pancreatic ductal adenocarcinoma. Cell Res 2019;29(9):725–38.

93. Xue M, Zhu Y, Jiang Y, Han L, Shi M, Su R et al. Schwann cells regulate tumor cells and cancer-associated fibroblasts in the pancreatic ductal adenocarcinoma microenvironment. Nat Commun 2023;14(1):1–18.

94. Gayoso A, Lopez R, Xing G, Boyeau P, Valiollah Pour Amiri V, Hong J et al. A Python library for probabilistic analysis of single-cell omics data. Nat Biotechnol 2022;40(2):163–6.

95. Sainburg T, McInnes L, Gentner TQ. Parametric UMAP embeddings for representation and semi-supervised learning. 2021;2025(Jun 17)

96. Xu C, Lopez R, Mehlman E, Regier J, Jordan MI, Yosef N. Probabilistic harmonization and annotation of single-cell transcriptomics data with deep generative models. Mol Syst Biol 2021;17(1):e9620.

97. Haghverdi L, Lun ATL, Morgan MD, Marioni JC. Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. Nat Biotechnol 2018;36(5):421–7.

98. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM et al. Comprehensive Integration of Single-Cell Data. Cell 2019;177(7):1888,1902.e21.

99. Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K et al. Fast, sensitive and accurate integration of single-cell data with Harmony. Nat Methods 2019;16(12):1289–96.

100. Lopez R, Regier J, Cole MB, Jordan MI, Yosef N. Deep generative modeling for single-cell transcriptomics. Nat Methods 2018;15(12):1053–8.

101. Rand WM. Objective Criteria for the Evaluation of Clustering Methods. Journal of the American Statistical Association 1971;66(336):846–50.

102. Luecken MD, Büttner M, Chaichoompu K, Danese A, Interlandi M, Mueller MF et al. Benchmarking atlas-level data integration in single-cell genomics. Nat Methods 2022;19(1):41–50.

103. Song Y, Miao Z, Brazma A, Papatheodorou I. Benchmarking strategies for cross-species integration of single-cell RNA sequencing data. Nat Commun 2023;14(1):6495.

104. Korotkevich G, Sukhov V, Budin N, Shpak B, Artyomov MN, Sergushichev A. Fast gene set enrichment analysis. 2021;:060012.

105. Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. Bioinformatics 2010;26(12):1572–3.

106. Blondel VD, Guillaume J, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. J Stat Mech 2008;2008(10):P10008.

107. Cui Zhou D, Jayasinghe RG, Chen S, Herndon JM, Iglesia MD, Navale P et al. Spatially restricted drivers and transitional cell populations cooperate with the microenvironment in untreated and chemo-resistant pancreatic cancer. Nat Genet 2022;54(9):1390–405.

108. Ma Y, Zhou X. Spatially informed cell-type deconvolution for spatial transcriptomics. Nat Biotechnol 2022;40(9):1349–59.

109. Anselin Luc. Local Indicators of Spatial Association—LISA. Geographical Analysis 1995;27(2):93–115.

110. Pham D, Tan X, Balderson B, Xu J, Grice LF, Yoon S et al. Robust mapping of spatiotemporal trajectories and cell–cell interactions in healthy and diseased tissues. Nat Commun 2023;14(1):7739.

111. Browaeys R, Gilis J, Sang-Aram C, Bleser PD, Hoste L, Tavernier S et al. MultiNicheNet: a flexible framework for differential cell-cell communication analysis from multi-sample multi-condition single-cell transcriptomics data. 2023;:2023.06.13.544751.

112. Jin S, Guerrero-Juarez CF, Zhang L, Chang I, Ramos R, Kuan C et al. Inference and analysis of cell-cell communication using CellChat. Nat Commun 2021;12(1):1088.

113. Raphael BJ, Hruban RH, Aguirre AJ, Moffitt RA, Yeh JJ, Stewart C et al. Integrated Genomic Characterization of Pancreatic Ductal Adenocarcinoma. Cancer Cell 2017;32(2):185,203.e13.

114. Colaprico A, Silva TC, Olsen C, Garofano L, Cava C, Garolini D et al. TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. Nucleic Acids Res 2016;44(8):e71.

115. Cao L, Huang C, Cui Zhou D, Hu Y, Lih TM, Savage SR et al. Proteogenomic characterization of pancreatic ductal adenocarcinoma. Cell 2021;184(19):5031,5052.e26.

116. Puleo F, Nicolle R, Blum Y, Cros J, Marisa L, Demetter P et al. Stratification of Pancreatic Ductal Adenocarcinomas Based on Tumor and Microenvironment Features. Gastroenterology 2018;155(6):1999,2013.e3.

117. Link JM, Eng JR, Pelz C, MacPherson-Hawthorne K, Worth PJ, Sivagnanam S et al. Ongoing replication stress tolerance and clonal T cell responses distinguish liver and lung recurrence and outcomes in pancreatic cancer. Nat Cancer 2025;6(1):123–44.

118. Diana A, Wang LM, D'Costa Z, Allen P, Azad A, Silva MA et al. Prognostic value, localization and correlation of PD-1/PD-L1, CD8 and FOXP3 with the desmoplastic stroma in pancreatic ductal adenocarcinoma. Oncotarget 2016;7(27):40992–1004.

119. Parikh AR, Szabolcs A, Allen JN, Clark JW, Wo JY, Raabe M et al. Radiation therapy enhances immunotherapy response in microsatellite stable colorectal and pancreatic adenocarcinoma in a phase II trial. Nat Cancer 2021;2(11):1124–35.

120. Costa M. Probabilistic interpretation of feedforward network outputs, with relationships to statistical prediction of ordinal quantities. Int J Neural Syst 1996;7(5):627–37.

121. Newman AM, Steen CB, Liu CL, Gentles AJ, Chaudhuri AA, Scherer F et al. Determining cell type abundance and expression from bulk tissues with digital cytometry. Nat Biotechnol 2019;37(7):773–82.

122. Hu M, Chikina M. InstaPrism: an R package for fast implementation of BayesPrism. Bioinformatics 2024;40(7):btae440.

123. Wang X, Park J, Susztak K, Zhang NR, Li M. Bulk tissue cell type deconvolution with multi-subject single-cell expression reference. Nat Commun 2019;10(1):380.

124. Jew B, Alvarez M, Rahmani E, Miao Z, Ko A, Garske KM et al. Accurate estimation of cell composition in bulk expression through robust integration of single-cell information. Nat Commun 2020;11(1):1971.

125. Gao R, Bai S, Henderson YC, Lin Y, Schalck A, Yan Y et al. Delineating copy number and clonal substructure in human tumors from single-cell transcriptomes. Nat Biotechnol 2021;39(5):599–608.

126. Xue R, Zhang Q, Cao Q, Kong R, Xiang X, Liu H et al. Liver tumour immune microenvironment subtypes and neutrophil heterogeneity. Nature 2022;612(7938):141–7.

127. Chakravarty D, Gao J, Phillips SM, Kundra R, Zhang H, Wang J et al. OncoKB: A Precision Oncology Knowledge Base. JCO precision oncology 2017;2017:10.1200/PO.17.00011.

128. Bärthel S, Falcomatà C, Rad R, Theis FJ, Saur D. Single-cell profiling to explore pancreatic cancer heterogeneity, plasticity and response to therapy. Nat Cancer 2023;4(4):454–67.

129. Sikkema L, Ramírez-Suástegui C, Strobl DC, Gillett TE, Zappia L, Madissoon E et al. An integrated cell atlas of the lung in health and disease. Nat Med 2023;29(6):1563–77.

130. Reed AD, Pensa S, Steif A, Stenning J, Kunz DJ, Porter LJ et al. A single-cell atlas enables mapping of homeostatic cellular shifts in the adult human breast. Nat Genet 2024;56(4):652–62.

131. Chijimatsu R, Kobayashi S, Takeda Y, Kitakaze M, Tatekawa S, Arao Y et al. Establishment of a reference single-cell RNA sequencing dataset for human pancreatic adenocarcinoma. iScience 2022;25(8):104659.

132. Zhang S, Fang W, Zhou S, Zhu D, Chen R, Gao X et al. Single cell transcriptomic analyses implicate an immunosuppressive tumor microenvironment in pancreatic cancer liver metastasis. Nat Commun 2023;14(1):5123.

133. Werba G, Weissinger D, Kawaler EA, Zhao E, Kalfakakou D, Dhara S et al. Single-cell RNA sequencing reveals the effects of chemotherapy on human pancreatic adenocarcinoma and its tumor microenvironment. Nat Commun 2023;14:797.

134. Sikkema L, Ramírez-Suástegui C, Strobl DC, Gillett TE, Zappia L, Madissoon E et al. An integrated cell atlas of the lung in health and disease. Nat Med 2023;29(6):1563–77.

135. Luecken MD, Theis FJ. Current best practices in single-cell RNA-seq analysis: a tutorial. Mol Syst Biol 2019;15(6):e8746.

136. Guimarães GR, Maklouf GR, Teixeira CE, de Oliveira Santos L, Tessarollo NG, de Toledo NE et al. Single-cell resolution characterization of myeloid-derived cell states with implication in cancer outcome. Nat Commun 2024;15(1):5694.

137. Wang H, Chen L, Qi L, Jiang N, Zhang Z, Guo H et al. A Single-Cell Atlas of Tumor-Infiltrating Immune Cells in Pancreatic Ductal Adenocarcinoma. Molecular & Cellular Proteomics 2022;21(8):100258.

138. Go S, Demetriou C, Valenzano G, Hughes S, Lanfredini S, Ferry H et al. Tissue-resident natural killer cells support survival in pancreatic cancer through promotion of cDC1-CD8 T activity. Elife 2024;13:RP92672.

139. Li H, Yang Z, Guo Q. Immune checkpoint inhibition for pancreatic ductal adenocarcinoma: limitations and prospects: a systematic review. Cell Communication and Signaling 2021;19(1):117.

140. Zhu Y, Knolhoff BL, Meyer MA, Nywening TM, West BL, Luo J et al. CSF1/CSF1R Blockade Reprograms Tumor-Infiltrating Macrophages and Improves Response to T Cell Checkpoint Immunotherapy in Pancreatic Cancer Models. Cancer Res 2014;74(18):5057–69.

141. Weeden CE, Gayevskiy V, Marceaux C, Batey D, Tan T, Yokote K et al. Early immune pressure initiated by tissue-resident memory T cells sculpts tumor evolution in non-small cell lung cancer. Cancer Cell 2023;41(5):837,852.e6.

142. Willemsen M, Tio D, Krebbers G, Kasiem FR, Jaspars EH, Matos TR et al. Presence of Skin Tissue-Resident Memory T Cells in Human Nonmalignant and Premalignant Melanocytic Skin Lesions and in Melanoma. Am J Dermatopathol 2022;44(6):416–23.

143. Zhang S, Fang W, Zhou S, Zhu D, Chen R, Gao X et al. Single cell transcriptomic analyses implicate an immunosuppressive tumor microenvironment in pancreatic cancer liver metastasis. Nat Commun 2023;14(1):5123.

144. Hegde S, Krisnawan VE, Herzog BH, Zuo C, Breden MA, Knolhoff BL et al. Dendritic cell paucity leads to dysfunctional immune surveillance in pancreatic cancer. Cancer Cell 2020;37(3):289,307.e9.

145. Katsuta E, Qi Q, Peng X, Hochwald SN, Yan L, Takabe K. Pancreatic adenocarcinomas with mature blood vessels have better overall survival. Sci Rep 2019;9(1):1310.

146. Li Y, Zheng Y, Huang J, Nie R, Wu Q, Zuo Z et al. CAF-macrophage crosstalk in tumour microenvironments governs the response to immune checkpoint blockade in gastric cancer peritoneal metastases. Gut 2025;74(3):350–63.

147. Dreschers S, Platen C, Ludwig A, Gille C, Köstlin N, Orlikowsky TW. Metalloproteinases TACE and MMP-9 Differentially Regulate Death Factors on Adult and Neonatal Monocytes After Infection with Escherichia coli. International Journal of Molecular Sciences 2019;20(6):1399.

148. Lu Z, Wang X, Feng J, Chai W, Wang W, Wang Q et al. Intratumoral CXCR4hi neutrophils display ferroptotic and immunosuppressive signatures in hepatoblastoma. Front Immunol 2024;15

149. Hwang WL, Jagadeesh KA, Guo JA, Hoffman HI, Yadollahpour P, Reeves JW et al. Single-nucleus and spatial transcriptome profiling of pancreatic cancer identifies multicellular dynamics associated with neoadjuvant treatment. Nat Genet 2022;54(8):1178–91.

150. Zhang X, Lan R, Liu Y, Pillarisetty VG, Li D, Zhao CL et al. Complement activation in tumor microenvironment after neoadjuvant therapy and its impact on pancreatic cancer outcomes. npj Precis Onc 2025;9(1):58.

151. Wang X, Guo L, Huang J, Jiang S, Li N, Mu H et al. Plasminogen Activator Inhibitor-1 Potentiates Neutrophil Infiltration and Tissue Injury in Colitis. Int J Biol Sci 2023;19(7):2132–49.

152. Araujo AM, Abaurrea A, Azcoaga P, López-Velazco JI, Manzano S, Rodriguez J et al. Stromal oncostatin M cytokine promotes breast cancer progression by reprogramming the tumor microenvironment. J Clin Invest 132(7):e148667.

153. Teng H, Wang T, Lin C, Tong Z, Cheng H, Wang H. Interferon Gamma Induces Higher Neutrophil Extracellular Traps Leading to Tumor-Killing Activity in Microsatellite Stable Colorectal Cancer. Mol Cancer Ther 2024;23(7):1043–56.

154. Grolleau-Julius A, Harning EK, Abernathy LM, Yung RL. Impaired Dendritic Cell Function in Aging Leads to Defective Antitumor Immunity. Cancer Research 2008;68(15):6341–9.

155. Hu C, Liu H, Pang B, Wu H, Lin X, Zhen Y et al. Supraphysiological estradiol promotes human T follicular helper cell differentiation and favours humoural immunity during in vitro fertilization. J Cell Mol Med 2021;25(14):6524–34.

156. Manoukian P, Dings M, Koster J, Medema J, Laarhoven Hv, Bijlsma M. P-168 Estrogen signaling shapes a tumor suppressive stroma in pancreatic cancer. Annals of Oncology 2023;34:S75.

157. Iribar H, Jaka A, Ormaechea N, Tuneu A, Izeta A, Gutiérrez-Rivera A. Does Schwann cell dedifferentiation originate dermal neurofibromas? Exp Dermatol 2016;25(11):901–3.

158. Kim H, Kim JY, Song CL, Jeong JE, Cho YS. Directly induced human Schwann cell precursors as a valuable source of Schwann cells. Stem Cell Res Ther 2020;11:257.

159. Gautam SK, Batra SK, Jain M. Molecular and metabolic regulation of immunosuppression in metastatic pancreatic ductal adenocarcinoma. Mol Cancer 2023;22:118.

160. Wu Y, Yang S, Ma J, Chen Z, Song G, Rao D et al. Spatiotemporal Immune Landscape of Colorectal Cancer Liver Metastasis at Single-Cell Level. Cancer Discov 2022;12(1):134–53.

161. Voissière A, Gomez-Roca C, Chabaud S, Rodriguez C, Nkodia A, Berthet J et al. The CSF-1R inhibitor Pexidartinib impacts dendritic cell differentiation through inhibition of FLT3 signaling and may antagonize the effect of durvalumab in patients with advanced cancer – results from a phase 1 study. 2023;:2023.02.15.23285939.

162. Halvorsen EC, Hamilton MJ, Young A, Wadsworth BJ, LePard NE, Lee HN et al. Maraviroc decreases CCL8-mediated migration of CCR5+ regulatory T cells and reduces metastatic tumor growth in the lungs. Oncoimmunology 2016;5(6):e1150398.

163. Hemmatazad H, Berger MD. CCR5 is a potential therapeutic target for cancer. Expert Opin Ther Targets 2021;25(4):311–27.

164. Davendra Sohal, Mai T. Duong, Renee Chang, Yuqing Xue, Devora Delman, Ignacio Garrido-Laguna, Sean J. Mulvihill, Kajsa Affolter, Mary Kay Washington, Muhammad Shaalan Beg, Andrea Wang-Gillam, James Lloyd Wade, Namita Gandhi, Syed A. Ahmad, Andrew M. Lowy, Elena Gabriela Chiorean, Katherine A Guthrie, Howard S. Hochster, Philip Agop Philip, and Gregory Lawrence Beatty. Immunologic predictors of therapeutic response to neoadjuvant

chemotherapy for pancreatic ductal adenocarcinoma (PDA) in SWOG S1505. | Journal of Clinical Oncology. Journal of Clinical Oncology 2021;39(3)

165. Jabbari N, Kenerson HL, Lausted C, Yan X, Meng C, Sullivan KM et al. Modulation of Immune Checkpoints by Chemotherapy in Human Colorectal Liver Metastases. Cell Rep Med 2020;1(9):100160.

166. Zhang S, Lv K, Liu Z, Zhao R, Li F. Fatty acid metabolism of immune cells: a new target of tumour immunotherapy. Cell Death Discov 2024;10(1):39.

167. Cheng C, Hu J, Mannan R, He T, Bhattacharyya R, Magnuson B et al. Targeting PIKfyve-driven lipid metabolism in pancreatic cancer. Nature 2025;642(8068):776–84.

168. Lotfollahi M, Naghipourfar M, Luecken MD, Khajavi M, Büttner M, Wagenstetter M et al. Mapping single-cell data to reference atlases by transfer learning. Nat Biotechnol 2022;40(1):121–30.

169. Binnewies M, Roberts EW, Kersten K, Chan V, Fearon DF, Merad M et al. Understanding the tumor immune microenvironment (TIME) for effective therapy. Nat Med 2018;24(5):541–50.

170. Xiao L, Shen Z, Pan Z, Qiu Y, Huang D, Liu Y et al. High-dimensional deconstruction of HNSC reveals clinically distinct cellular states and ecosystems that are associated with prognosis and therapy response. Journal of Translational Medicine 2025;23(1):254.

171. Xu X, Li R, Mo O, Liu K, Li J, Hao P. Cell-type deconvolution for bulk RNA-seq data using single-cell reference: a comparative analysis and recommendation guideline. Brief Bioinform 2025;26(1):bbaf031.

172. Yang Z, Algesheimer R, Tessone CJ. A Comparative Analysis of Community Detection Algorithms on Artificial Networks. Sci Rep 2016;6(1):30750.

173. Schubert M, Klinger B, Klünemann M, Sieber A, Uhlitz F, Sauer S et al. Perturbation-response genes reveal signaling footprints in cancer gene expression. Nat Commun 2018;9(1):20.

174. Li H, Zhou J, Li Z, Chen S, Liao X, Zhang B et al. A comprehensive benchmarking with practical guidelines for cellular deconvolution of spatial transcriptomics. Nat Commun 2023;14(1):1548.

175. Shi Q, Chen Y, Li Y, Qin S, Yang Y, Gao Y et al. Cross-tissue multicellular coordination and its rewiring in cancer. Nature 2025;643(8071):529–38.

176. Xie Z, Niu L, Zheng G, Du K, Dai S, Li R et al. Single-cell analysis unveils activation of mast cells in colorectal cancer microenvironment. Cell & Bioscience 2023;13(1):217.

177. Le K, Sun J, Ghaemmaghami J, Smith MR, Ip WKE, Phillips T et al. Blockade of CCR1 induces a phenotypic shift in macrophages and triggers a favorable antilymphoma activity. Blood Adv 2023;7(15):3952–67.

178. Wen Z, Liu T, Xu X, Acharya N, Shen Z, Lu Y et al. Interleukin-16 enhances anti-tumor immune responses by establishing a Th1 cell-macrophage crosstalk through reprogramming glutamine metabolism in mice. Nat Commun 2025;16(1):2362.

179. Mohamed AH, Obeid RA, Fadhil AA, Amir AA, Adhab ZH, Jabouri EA et al. BTLA and HVEM: Emerging players in the tumor microenvironment and cancer progression. Cytokine 2023;172:156412.

180. Filippou PS, Karagiannis GS, Constantinidou A. Midkine (MDK) growth factor: a key player in cancer progression and a promising therapeutic target. Oncogene 2020;39(10):2040–54.

181. Rossi Sebastiano M, Pozzato C, Saliakoura M, Yang Z, Peng R, Galiè M et al. ACSL3–PAI-1 signaling axis mediates tumor-stroma cross-talk promoting pancreatic cancer progression. Sci Adv 2020;6(44):eabb9200.

182. Kubala MH, Punj V, Placencio-Hickok VR, Fang H, Fernandez GE, Sposto R et al. Plasminogen Activator Inhibitor-1 Promotes the Recruitment and Polarization of Macrophages in Cancer. Cell Reports 2018;25(8):2177,2191.e7.

183. Pandey P, Khan F, Upadhyay TK, Seungjoon M, Park MN, Kim B. New insights about the PDGF/PDGFR signaling pathway as a promising target to develop cancer therapeutic strategies. Biomedicine & Pharmacotherapy 2023;161:114491.

184. Sivakumar S, Jainarayanan A, Arbe-Barnes E, Sharma PK, Leathlobhair MN, Amin S et al. Distinct immune cell infiltration patterns in pancreatic ductal adenocarcinoma (PDAC) exhibit divergent immune cell selection and immunosuppressive mechanisms. Nat Commun 2025;16(1):1397.

185. Royal RE, Levy C, Turner K, Mathur A, Hughes M, Kammula US et al. Phase 2 Trial of Single Agent Ipilimumab (Anti-CTLA-4) for Locally Advanced or Metastatic Pancreatic Adenocarcinoma. Journal of Immunotherapy 2010;33(8):828.

186. O'Reilly EM, Oh D, Dhani N, Renouf DJ, Lee MA, Sun W et al. Durvalumab With or Without Tremelimumab for Patients With Metastatic Pancreatic Ductal Adenocarcinoma: A Phase 2 Randomized Clinical Trial. JAMA Oncol 2019;5(10):1431–8.

187. Muller M, Haghnejad V, Schaefer M, Gauchotte G, Caron B, Peyrin-Biroulet L et al. The Immune Landscape of Human Pancreatic Ductal Carcinoma: Key Players, Clinical Implications, and Challenges. Cancers (Basel) 2022;14(4):995.

188. Qi J, Sun H, Zhang Y, Wang Z, Xun Z, Li Z et al. Single-cell and spatial analysis reveal interaction of FAP+ fibroblasts and SPP1+ macrophages in colorectal cancer. Nat Commun 2022;13(1):1742.

189. Xun Z, Ding X, Zhang Y, Zhang B, Lai S, Zou D et al. Reconstruction of the tumor spatial microenvironment along the malignant-boundary-nonmalignant axis. Nat Commun 2023;14(1):933.

190. Fujimura T, Yoshino K, Kato H, Fukushima S, Ishizuki S, Otsuka A et al. A phase II multicentre study of plasminogen activator inhibitor-1 inhibitor (TM5614) plus nivolumab for treating anti-programmed cell death 1 antibody-refractory malignant melanoma: TM5614-MM trial. Br J Dermatol 2024;191(5):691–7.

191. Sturm N, Ettrich TJ, Perkhofer L. The Impact of Biomarkers in Pancreatic Ductal Adenocarcinoma on Diagnosis, Surveillance and Therapy. Cancers (Basel) 2022;14(1):217.

192. Daamen LA, Groot VP, Intven MPW, Besselink MG, Busch OR, Koerkamp BG et al. Postoperative surveillance of pancreatic cancer patients. European Journal of Surgical Oncology 2019;45(10):1770–7.

193. van Oosten AF, Groot VP, Dorland G, Burkhart RA, Wolfgang CL, van Santvoort HC et al. Dynamics of Serum CA19-9 in Patients Undergoing Pancreatic Cancer Resection. Ann Surg 2024;279(3):493–500.

194. Alvikas J, Hamed A, Tirukkovalur N, Ceuppens S, Tcharni A, Qian J et al. Survival prediction for CA19-9 non-producers with resected pancreatic cancer. HPB 2025;27(9):1185–93.

195. Lin M, Huang J, Yu H. Elevated serum level of carbohydrate antigen 19-9 in benign biliary stricture diseases can reduce its value as a tumor marker. Int J Clin Exp Med 2014;7(3):744–50.

196. Haab BB, Huang Y, Balasenthil S, Partyka K, Tang H, Anderson M et al. Definitive Characterization of CA 19-9 in Resectable Pancreatic Cancer Using a Reference Set of Serum and Plasma Specimens. PLOS ONE 2015;10(10):e0139049.

197. Lansbergen MF, Dings MPG, Manoukian P, Fariña A, Waasdorp C, Hooijer GKJ et al. Transcriptome-based classification to predict FOLFIRINOX response in a real-world metastatic pancreatic cancer cohort. Translational Research 2024;273:137–47.

198. Aung KL, Fischer SE, Denroche RE, Jang G, Dodd A, Creighton S et al. Genomics-Driven Precision Medicine for Advanced Pancreatic Cancer: Early Results from the COMPASS Trial. Clin Cancer Res 2018;24(6):1344–54.

199. Bill R, Wirapati P, Messemaker M, Roh W, Zitti B, Duval F et al. CXCL9:SPP1 macrophage polarity identifies a network of cellular programs that control human cancers. Science 2023;381(6657):515–24.

200. Xiang Z, Hu T, Wang Y, Wang H, Xu L, Cui N. Neutrophil–lymphocyte ratio (NLR) was associated with prognosis and immunomodulatory in patients with pancreatic ductal adenocarcinoma (PDAC). Biosci Rep 2020;40(6):BSR20201190.

201. Chu T, Wang Z, Pe’er D, Danko CG. Cell type and gene expression deconvolution with BayesPrism enables Bayesian integrative analysis across bulk and single-cell RNA sequencing in oncology. Nat Cancer 2022;3(4):505–17.

202. Hiraoka N, Ino Y, Yamazaki-Itoh R, Kanai Y, Kosuge T, Shimada K. Intratumoral tertiary lymphoid organ is a favourable prognosticator in patients with pancreatic cancer. Br J Cancer 2015;112(11):1782–90.

203. Bagaev A, Kotlov N, Nomie K, Svekolkin V, Gafurov A, Isaeva O et al. Conserved pan-cancer microenvironment subtypes predict response to immunotherapy. Cancer Cell 2021;39(6):845,865.e7.

204. Thorsson V, Gibbs DL, Brown SD, Wolf D, Bortone DS, Ou Yang T et al. The Immune Landscape of Cancer. Immunity 2018;48(4):812,830.e14.

205. Patel RB, Hernandez R, Carlson P, Grudzinski J, Bates AM, Jagodinsky JC et al. Low-dose targeted radionuclide therapy renders immunologically cold tumors responsive to immune checkpoint blockade. Science Translational Medicine 2021;13(602):eabb3631.

206. Xing R, Mei J, Zuo Z, Zou H, Yu X, Xu J et al. Enhanced formation of tertiary lymphoid structures shapes the anti-tumor microenvironment in hepatocellular carcinoma after FOLFOX-HAIC therapy. Cell Rep Med 2025;:102298.

207. Sezgin Y, Karhan O, Aldemir MN, Ürün M, Erçek BM, Urakcı Z et al. Efficacy of gemcitabine plus nab-paclitaxel in second-line treatment of metastatic pancreatic cancer. Sci Rep 2025;15(1):11675.

208. Zhang J, Song J, Tang S, Zhao Y, Wang L, Luo Y et al. Multi-omics analysis reveals the chemoresistance mechanism of proliferating tissue-resident macrophages in PDAC via metabolic adaptation. Cell Reports 2023;42(6)

209. Ullman NA, Burchard PR, Dunne RF, Linehan DC. Immunologic Strategies in Pancreatic Cancer: Making Cold Tumors Hot. J Clin Oncol 2022;40(24):2789–805.

210. Riaz N, Havel JJ, Makarov V, Desrichard A, Urba WJ, Sims JS et al. Tumor and Microenvironment Evolution during Immunotherapy with Nivolumab. Cell 2017;171(4):934,949.e16.

211. Yang F, Shaibu Z, Liu Q, Zhu W. Cytokine profiles as predictive biomarkers for treatment outcomes in advanced gastric cancer patients undergoing PD-1 blockade immunochemotherapy: a meta-analysis. Clin Exp Med 2025;25(1):136.

212. Larsen BM, Kannan M, Langer LF, Leibowitz BD, Bentaieb A, Cancino A et al. A pan-cancer organoid platform for precision medicine. Cell Rep 2021;36(4):109429.

213. Hughes D, Evans A, Go S, Eyres M, Pan L, Mukherjee S et al. Development of human pancreatic cancer avatars as a model for dynamic immune landscape profiling and personalized therapy. Sci Adv 10(27):eadm9071.

214. Wlodarczyk B, Gasiorowska A, Malecka-Panas E. The Role of Insulin-like Growth Factor (IGF) Axis in Early Diagnosis of Pancreatic Adenocarcinoma (PDAC). J Clin Gastroenterol 2018;52(7):569–72.

215. Wlodarczyk B, Durko L, Wlodarczyk P, Talar-Wojnarowska R, Malecka-Wojciesko E. CA 19-9 but Not IGF-1/IGFBP-2 Is a Useful Biomarker for Pancreatic Ductal Adenocarcinoma (PDAC) and Chronic Pancreatitis (CP) Differentiation. Journal of Clinical Medicine 2023;12(12):4050.

216. Park WG, Li L, Appana S, Wei W, Stello K, Andersen DK et al. Unique Circulating Immune Signatures for Recurrent Acute Pancreatitis, Chronic Pancreatitis and Pancreatic Cancer: A Pilot Study of These Conditions with and without Diabetes. Pancreatology 2020;20(1):51–9.

217. van der Sijde F, van Dam JL, Groot Koerkamp B, Haberkorn BCM, Homs MYV, Mathijssen D et al. Treatment Response and Conditional Survival in Advanced Pancreatic Cancer Patients Treated with FOLFIRINOX: A Multicenter Cohort Study. J Oncol 2022;2022:8549487.

218. Eijck CWFv, Koning Wd, Sijde Fvd, Moskie M, Koerkamp BG, Homs MYV et al. A multigene circulating biomarker to predict the lack of FOLFIRINOX response after a single cycle in patients with pancreatic ductal adenocarcinoma. European Journal of Cancer 2023;181:119–34.

219. Ullman NA, Burchard PR, Dunne RF, Linehan DC. Immunologic Strategies in Pancreatic Cancer: Making Cold Tumors Hot. J Clin Oncol 2022;40(24):2789–805.

220. Zhang E, Ma Y, Liu Z, Zhang J, Liu W, Chen Y et al. Prognostic implications and characterization of tumor-associated tertiary lymphoid structures genes in pancreatic cancer. Journal of Translational Medicine 2025;23(1):301.

221. Galon J, Bruni D. Approaches to treat immune hot, altered and cold tumours with combination immunotherapies. Nat Rev Drug Discov 2019;18(3):197–218.

222. Subramanian A, Nemat-Gorgani N, Ellis-Caleo TJ, van IJzendoorn DGP, Sears TJ, Somani A et al. Sarcoma microenvironment cell states and ecosystems are associated with prognosis and predict response to immunotherapy. Nat Cancer 2024;5(4):642–58.

223. Morizane C, Ueno M, Ikeda M, Sudo K, Hirashima Y, Kuroda M et al. A Phase 2 study of nivolumab in combination with modified FOLFIRINOX for metastatic pancreatic cancer. BJC Rep 2024;2:3.

224. Gavish A, Tyler M, Greenwald AC, Hoefflin R, Simkin D, Tschernichovsky R et al. Hallmarks of transcriptional intratumour heterogeneity across a thousand tumours. Nature 2023;618(7965):598–606.

225. Fan J, Lu F, Qin T, Peng W, Zhuang X, Li Y et al. Multiomic analysis of cervical squamous cell carcinoma identifies cellular ecosystems with biological and clinical relevance. Nat Genet 2023;55(12):2175–88.

226. Weaver BA, Cleveland DW. Does aneuploidy cause cancer? Current Opinion in Cell Biology 2006;18(6):658–67.

227. Song M, Ma S, Wang G, Wang Y, Yang Z, Xie B et al. Benchmarking copy number aberrations inference tools using single-cell multi-omics datasets. Brief Bioinform 2025;26(2):bbaf076.

228. Harada T, Chelala C, Bhakta V, Chaplin T, Caulee K, Baril P et al. Genome-wide DNA copy number analysis in pancreatic cancer using high-density single nucleotide polymorphism arrays. Oncogene 2008;27(13):1951–60.

229. Zhao K, Vos J, Lam S, Boe LA, Muldoon D, Han CY et al. Longitudinal and multisite sampling reveals mutational and copy number evolution in tumors during metastatic dissemination. Nat Genet 2025;57(6):1504–11.

230. Liu W, Rodgers GP. Olfactomedin 4 Is Not a Precise Marker for Human Intestinal Stem Cells, But Is Involved in Intestinal Carcinogenesis. Gastroenterology 2022;162(4):1001–4.

231. Jiang Y, Liao C, Lai J, Peng Y, Chen Q, Zheng X. KRT7 promotes pancreatic cancer metastasis by remodeling the extracellular matrix niche through FGF2-fibroblast crosstalk. Sci Rep 2025;15(1):6951.

232. Khantakova D, Brioschi S, Molgora M. Exploring the Impact of TREM2 in Tumor-Associated Macrophages. Vaccines (Basel) 2022;10(6):943.

233. Chao MP, Weissman IL, Majeti R. The CD47-SIRPα Pathway in Cancer Immune Evasion and Potential Therapeutic Implications. Curr Opin Immunol 2012;24(2):225–32.

234. Kajiwara T, Tanaka T, Kukita K, Kutomi G, Saito K, Okuya K et al. Hypoxia augments MHC class I antigen presentation via facilitation of ERO1-α-mediated oxidative folding in murine tumor cells. Eur J Immunol 2016;46(12):2842–51.

235. Csiszar A, Ahmad M, Smith KE, Labinskyy N, Gao Q, Kaley G et al. Bone Morphogenetic Protein-2 Induces Proinflammatory Endothelial Phenotype. Am J Pathol 2006;168(2):629–38.

236. Foley K, Rucki AA, Xiao Q, Zhou D, Leubner A, Mo G et al. Semaphorin 3D autocrine signaling mediates the metastatic role of annexin A2 in pancreatic cancer. Sci Signal 2015;8(388):ra77.

237. Li J, Peng L, Chen Q, Ye Z, Zhao T, Hou S et al. Integrin β1 in Pancreatic Cancer: Expressions, Functions, and Clinical Implications. Cancers 2022;14(14):3377.

238. Jerby-Arnon L, Tooley K, Escobar G, Dandekar G, Madi A, Goldschmidt E et al. Pan-cancer mapping of single T cell profiles reveals a TCF1:CXCR6-CXCL16 regulatory axis essential for effective anti-tumor immunity. 2021;:2021.10.31.466532.

239. Somerville TDD, Xu Y, Miyabayashi K, Tiriac H, Cleary CR, Maia-Silva D et al. TP63-Mediated Enhancer Reprogramming Drives the Squamous Subtype of Pancreatic Ductal Adenocarcinoma. Cell Reports 2018;25(7):1741.

240. Wang J, Xu L, Liu C, Huang T, Liang C, Fan Y. Identifying the role of apolipoprotein A-I in prostate cancer. Asian J Androl 2021;23(4):400.

241. Oh K, Yoo YJ, Torre-Healy LA, Rao M, Fassler D, Wang P et al. Coordinated single-cell tumor microenvironment dynamics reinforce pancreatic cancer subtype. Nat Commun 2023;14(1):5226.

242. Bahmed K, Henry C, Holliday M, Redzic J, Ciobanu M, Zhang F et al. Extracellular cyclophilin-A stimulates ERK1/2 phosphorylation in a cell-dependent manner but broadly stimulates nuclear factor kappa B. Cancer Cell International 2012;12(1):19.

243. Zhang L, Song L, Xu Y, Xu Y, Zheng M, Zhang P et al. Midkine promotes breast cancer cell proliferation and migration by upregulating NR3C1 expression and activating the NF-κB pathway. Mol Biol Rep 2022;49(4):2953–61.

244. George B, Kudryashova O, Kravets A, Thalji S, Malarkannan S, Kurzrock R et al. Transcriptomic-Based Microenvironment Classification Reveals Precision Medicine Strategies for Pancreatic Ductal Adenocarcinoma. Gastroenterology 2024;166(5):859,871.e3.

245. Martínez-Jiménez F, Priestley P, Shale C, Baber J, Rozemuller E, Cuppen E. Genetic immune escape landscape in primary and metastatic cancer. Nat Genet 2023;55(5):820–31.

246. Karamitopoulou E, Andreou A, Wenning AS, Gloor B, Perren A. High tumor mutational burden (TMB) identifies a microsatellite stable pancreatic cancer subset with prolonged survival and strong anti-tumor immunity. European Journal of Cancer 2022;169:64–73.

247. Wang H, Lin Y, Hsu C, Chao Y, Hou Y, Chiu T et al. Pancreatic stellate cells activated by mutant KRAS-mediated PAI-1 upregulation foster pancreatic cancer progression via IL-8. Theranostics 2019;9(24):7168–83.

248. Maddipati R, Norgard RJ, Baslan T, Rathi KS, Zhang A, Saeid A et al. MYC Levels Regulate Metastatic Heterogeneity in Pancreatic Adenocarcinoma. Cancer Discov 2021;12(2):542.

249. Ngan E, Kiepas A, Brown CM, Siegel PM. Emerging roles for LPP in metastatic cancer progression. J Cell Commun Signal 2018;12(1):143–56.

250. Xu P, Zhang Q, Zhai J, Chen P, Deng X, Miao L et al. APOA1 promotes tumor proliferation and migration and may be a potential pan-cancer biomarker and immunotherapy target. Translational Oncology 2025;55

251. Wang S, Yan W, Kong L, Zuo S, Wu J, Zhu C et al. Oncolytic viruses engineered to enforce cholesterol efflux restore tumor-associated macrophage phagocytosis and anti-tumor immunity in glioblastoma. Nat Commun 2023;14(1):4367.

252. Goossens P, Rodriguez-Vita J, Etzerodt A, Masse M, Rastoin O, Gouirand V et al. Membrane Cholesterol Efflux Drives Tumor-Associated Macrophage Reprogramming and Tumor Progression. Cell Metabolism 2019;29(6):1376,1389.e4.

253. Luecken MD, Büttner M, Chaichoompu K, Danese A, Interlandi M, Mueller MF et al. Benchmarking atlas-level data integration in single-cell genomics. Nat Methods 2022;19(1):41–50.

254. Wu Y, Yang S, Ma J, Chen Z, Song G, Rao D et al. Spatiotemporal Immune Landscape of Colorectal Cancer Liver Metastasis at Single-Cell Level. Cancer Discov 2022;12(1):134–53.

255. Nguyen H, Nguyen H, Tran D, Draghici S, Nguyen T. Fourteen years of cellular deconvolution: methodology, applications, technical evaluation and outstanding challenges. Nucleic Acids Res 2024;52(9):4761–83.

256. Khaliq AM, Rajamohan M, Saeed O, Mansouri K, Adil A, Zhang C et al. Spatial transcriptomic analysis of primary and metastatic pancreatic cancers highlights tumor microenvironmental heterogeneity. Nat Genet 2024;56(11):2455–65.

257. Özdemir BC, Pentcheva-Hoang T, Carstens JL, Zheng X, Wu C, Simpson TR et al. Depletion of Carcinoma-Associated Fibroblasts and Fibrosis Induces Immunosuppression and Accelerates Pancreas Cancer with Reduced Survival. Cancer Cell 2014;25(6):719–34.

258. Janssen QP, Dam JLv, Bekkum MLv, Bonsing BA, Bos H, Bosscha KP et al. Neoadjuvant FOLFIRINOX versus neoadjuvant gemcitabine-based chemoradiotherapy in resectable and borderline resectable pancreatic cancer (PREOPANC-2): a multicentre, open-label, phase 3 randomised trial. The Lancet Oncology 2025;26(10):1346–56.

259. Eshmuminov D, Aminjonov B, Palm RF, Malleo G, Schmocker RK, Abdallah R et al. FOLFIRINOX or Gemcitabine-based Chemotherapy for Borderline Resectable and Locally Advanced Pancreatic Cancer: A Multi-institutional, Patient-Level, Meta-analysis and Systematic Review. Ann Surg Oncol 2023;30(7):4417–28.

260. Cui X, Gu X, Li D, Wu P, Sun N, Zhang C et al. Tertiary lymphoid structures as a biomarker in immunotherapy and beyond: Advancing towards clinical application. Cancer Letters 2025;613:217491.

261. Vanhersecke L, Brunet M, Guégan J, Rey C, Bougouin A, Cousin S et al. Mature tertiary lymphoid structures predict immune checkpoint inhibitor efficacy in solid tumors independently of PD-L1 expression. Nat Cancer 2021;2(8):794–802.

262. An Y, Sun J, Xu M, Xu J, Ma S, Liu C et al. Tertiary lymphoid structure patterns aid in identification of tumor microenvironment infiltration and selection of therapeutic agents in bladder cancer. Front Immunol 2022;13

263. Li H, Zhang M, Zhang B, Lin W, Li S, Xiong D et al. Mature tertiary lymphoid structures evoke intra-tumoral T and B cell responses via progenitor exhausted CD4+ T cells in head and neck cancer. Nat Commun 2025;16(1):4228.

264. Luca BA, Steen CB, Matusiak M, Azizi A, Varma S, Zhu C et al. Atlas of clinically distinct cell states and ecosystems across human solid tumors. Cell 2021;184(21):5482,5496.e28.

265. Chu T, Zhu G, Tang Z, Qu W, Yang R, Pan H et al. Metabolism archetype cancer cells induce protumor TREM2+ macrophages via oxLDL-mediated metabolic interplay in hepatocellular carcinoma. Nat Commun 2025;16(1):6770.

266. Xia X, Zhou Z, Zheng X, Tu C, Liu H, Hu Z et al. APOE deficiency triggers anti-tumour activity of macrophages in liver cancer. Cancer Gene Ther 2025;32(9):949–62.

267. Li H, van der Leun AM, Yofe I, Lubling Y, Gelbard-Solodkin D, van Akkooi ACJ et al. Dysfunctional CD8 T Cells Form a Proliferative, Dynamically Regulated Compartment within Human Melanoma. Cell 2019;176(4):775,789.e18.

268. Cheng D, Qiu K, Rao Y, Mao M, Li L, Wang Y et al. Proliferative exhausted CD8+ T cells exacerbate long-lasting anti-tumor effects in human papillomavirus-positive head and neck squamous cell carcinoma. eLife 2023;12:e82705.

269. Storrs EP, Chati P, Usmani A, Sloan I, Krasnick BA, Babbra R et al. High-dimensional deconstruction of pancreatic cancer identifies tumor microenvironmental and developmental stemness features that predict survival. npj Precis Onc 2023;7(1):105.

270. Detlefsen S, Boldt HB, Burton M, Thomsen MM, Rasmussen LG, Ørbeck SV et al. High overall copy number variation burden by genome-wide methylation profiling holds negative prognostic value in surgically treated pancreatic ductal adenocarcinoma. Human Pathology 2023;142:68–80.

271. Smeby J, Sveen A, Bergsland CH, Eilertsen IA, Danielsen SA, Eide PW et al. Exploratory analyses of consensus molecular subtype-dependent associations of TP53 mutations with immunomodulation and prognosis in colorectal cancer. ESMO Open 2019;4(3)

272. Muñoz-Fontela C, Mandinova A, Aaronson SA, Lee SW. Emerging roles of p53 and other tumour-suppressor genes in immune regulation. Nat Rev Immunol 2016;16(12):741–50.

273. Takaoka A, Hayakawa S, Yanai H, Stoiber D, Negishi H, Kikuchi H et al. Integration of interferon-α/β signalling to p53 responses in tumour suppression and antiviral defence. Nature 2003;424(6948):516–23.

274. Patterson MT, Firulyova MM, Xu Y, Hillman H, Bishop C, Zhu A et al. Trem2 promotes foamy macrophage lipid uptake and survival in atherosclerosis. Nat Cardiovasc Res 2023;2(11):1015–31.

275. Casey SC, Tong L, Li Y, Do R, Walz S, Fitzgerald KN et al. MYC Regulates the Anti-Tumor Immune Response through CD47 and PD-L1. Science 2016;352(6282):227–31.

276. Sinha S, Borcherding N, Renavikar PS, Crawford MP, Tsalikian E, Tansey M et al. An autoimmune disease risk SNP, rs2281808, in SIRPG is associated with reduced expression of SIRPγ and heightened effector state in human CD8 T-cells. Sci Rep 2018;8(1):15440.

277. Marguerie F, Saifi MA, Geary B, Barnes D, Jonsson AH, Ho I-. Enhancement of activation-induced T cell proliferation by SIRPG in a CD47-independent manner. 2025;:2025.05.01.651731.

278. Zhou T, Wu Y, Li S, Qiu X, Liu E, Xie Z et al. Multi-omic analysis of gallbladder cancer identifies distinct tumor microenvironments associated with disease progression. Nat Genet 2025;57(8):1935–49.

279. Chu Y, Dai E, Li Y, Han G, Pei G, Ingram DR et al. Pan-cancer T cell atlas links a cellular stress response state to immunotherapy resistance. Nat Med 2023;29(6):1550–62.